Helsinki University Biomedical Dissertation no. 53/2018

Novel Approaches for Pluripotent Reprogramming

Jere Weltner

Research Programs Unit, Molecular Neurology
Biomedicum Stem Cell Center
Faculty of Medicine
Doctoral Programme in Biomedicine
University of Helsinki
Helsinki, Finland

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Medicine of the University of Helsinki, for public examination in lecture room 1, Biomedicum 1, on Aug 24th 2018, at 12.

Helsinki 2018

Supervised by

Professor Timo Otonkoski
Research Programs Unit, Molecular Neurology,
Biomedicum Stem Cell Center, University of Helsinki and
Children's Hospital, Helsinki University Central Hospital
Helsinki, Finland

and

Docent Ras Trokovic
Research Programs Unit, Molecular Neurology,
Biomedicum Stem Cell Center, University of Helsinki
Helsinki, Finland

Reviewed by

Associate professor Noora Kotaja
Institute of Biomedicine, University of Turku
Turku, Finland

and

Docent Einari Niskanen
Institute of Biomedicine, University of Eastern Finland
Kuopio, Finland

Opponent

Doctor Jacob Hanna
Department of Molecular Genetics, Weizmann Institute of Science
Rehovot, Israel

The Faculty of Medicine uses the Urkund system (plagiarism recognition) to examine all doctoral dissertations.

ISBN 978-951-51-4423-2 (paperback)
ISBN 978-951-51-4424-9 (PDF)
ISSN 2342-3161 (print)
ISSN 2342-317X (online)

Dissertationes Scholae Doctoralis Ad Sanitatem Investigandam Universitatis Helsinkiensis

Unigrafia
Helsinki 2018

1. Table of Contents

2. Abstract

Somatic cells can be reprogrammed to pluripotent state by ectopic expression of a defined set of transcription factors. These induced pluripotent stem cells (iPSC) hold great potential for biomedical applications, such as disease modelling, drug discovery and cell therapies. The derivation of iPSCs is a complex multistep process that can commonly result in inefficient or incomplete conversion of the cells. The reprogramming efficiency and the quality of the reprogrammed cells can be affected by various components of the reprogramming method, including reprogramming vectors, starting cell populations and the choice of reprogramming factors. The aim of this thesis was to explore novel approaches for improving the pluripotent reprogramming outcome.

The particular aims of this thesis were to investigate the use of recombinant Adeno-associated virus (rAAV) as a gene transfer vector for cellular reprogramming, characterization of the effects of old donor age and long term passaging on the reprogramming of fibroblasts, and development of reprogramming methods based on CRISPR/Cas9-mediated activation of endogenous reprogramming factors.

In this study, rAAV-mediated transduction of mouse embryonic fibroblasts with OCT4, SOX2, KLF4 and C-MYC was found to successfully induce reprogramming to pluripotency. Unlike initially expected, the AAV vectors were integrated with high efficiency into the host genome during the reprogramming process, resulting in all analyzed iPSCs containing vector integrations. The high frequency of vector integration may limit the use of rAAV for generation of genetically intact iPSCs, however this vector may be useful for more specialized reprogramming applications where vector integration may be desirable.

Both donor age and passage number of human fibroblasts can affect the reprograming efficiency. In this study, we have characterized the reprogramming of 11 skin fibroblast lines from various ages and passages. Both donor age and the culture time of the donor fibroblasts correlated with reduction in pluripotent reprogramming efficiency. This effect was found to be associated with upregulation of cellular P21 expression and reduction in cell proliferation. Downregulation of P21 expression by siRNA treatment was able to promote reprogramming of late passage senescent fibroblasts. Therefore, this study demonstrated that the inhibition of P21 can be used as a mean to improve the reprogramming efficiency of late passage senescent donor fibroblasts.

To promote activation of endogenous pluripotency genes in reprogramming, we developed a gene activation system based on CRISPR/Cas9. Using this system, endogenous reprogramming factors can be activated by targeting the gene promoter with a deactivated Cas9 protein (dCas9) fused to a transactivation domain. This process was shown to be temporally controllable by fusing the dCas9 with a chemically inducible degradation domain. By optimizing the reprogramming factor guide composition, CRISPR/Cas9-mediated gene activation

(CRISPRa) could be used to derive iPSCs reprogrammed fully by targeted activation of endogenous genes. The efficient reprogramming of somatic cells by CRISPRa was found to be dependent on the inclusion of additional guides targeting an embryonic genome activation enriched Alu-motif (EEA-motif). Due to the direct targeting of endogenous loci and the high multiplexing capacity of CRISPRa, the reprogramming approach has a high potential for mediating comprehensive and specific reprogramming. This system has also potential uses for elucidating the function of endogenous gene regulatory elements, such as the EEA-motif or other endogenous transposable elements, on the reprogramming process. The CRISPRa-based gene activation tools developed in this study thus provide a powerful new way of affecting cellular reprogramming.

Overall, this thesis provides a number of novel tools and insights into the pluripotent reprogramming process. The results of this work can be used to develop more robust reprogramming methods and to improve the quality of reprogrammed cells.

3. List of Original Publications

This thesis work is based on the following articles (**I-IV**), referred in the text by their Roman numerals.

I. **Weltner J**, Anisimov A, Alitalo K, Otonkoski T, Trokovic R. *Induced pluripotent stem cell clones reprogrammed via recombinant adeno-associated virus-mediated transduction contain integrated vector sequences*. J Virol. 2012 Apr;86(8):4463-7. DOI: 10.1128/JVI.06302-11

II. Trokovic R, **Weltner J**, Noisa P, Raivio T, Otonkoski T. *Combined negative effect of donor age and time in culture on the reprogramming efficiency into induced pluripotent stem cells*. Stem Cell Res. 2015 Jul;15(1):254-62. DOI: 10.1016/j.scr.2015.06.001.

III. Balboa D, **Weltner J**, Eurola S, Trokovic R, Wartiovaara K, Otonkoski T. *Conditionally Stabilized dCas9 Activator for Controlling Gene Expression in Human Cell Reprogramming and Differentiation*. Stem Cell Reports. 2015 Sep 8;5(3):448-59. DOI: 10.1016/j.stemcr.2015.08.001. This publication is included in the PhD thesis of Diego Balboa, 2018, University of Helsinki, Finland.

IV. **Weltner J**, Balboa D, Katayama S, Bespalov M, Krjutškov K, Jouhilahti E M, Trokovic R, Kere J, Otonkoski T. *Human pluripotent reprogramming with CRISPR activators*. Nature Communications. 2018 Jul 6;9(1):2643. DOI: 10.1038/s41467-018-05067-x

4. Abbreviations

5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
AAV	Adeno Associated Virus
AID	Activation Induced Cytidine Deaminase
alt-NHEJ	Alternative Non-Homologous End Joining
BMP	Bone Morphogenetic Protein
cAMP	Cyclic Adenosine Monophosphate
Cas9	CRISPR-Associated Protein 9
CPP	Cell Penetrating Peptide
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeat
crRNA	CRISPR RNA
CTCF	CCCTC-Binding Factor
dCas9	Dead Cas9 (catalytically inactivated)
DMR	Differentially Methylated Region
DNA	Deoxyribonucleic Acid
DNMT	DNA Methyltransferase
DOX	Doxycycline
DSB	Double Strand Break
EB	Embryoid Body
EEA-motif	EGA Enriched Alu-motif
EGA	Embryonic Genome Activation
EMT	Epithelial To Mesenchymal Transition
EpiSC	Epiblast Stem Cell
ERV	Endogenous Retrovirus
ESC	Embryonic Stem Cell
FGF	Fibroblast Growth Factor
gRNA	Guide RNA
HDAC	Histone Deacetylase
HDR	Homology Directed Repair
HFF	Human Foreskin Fibroblast
HMG	High Mobility Group
ICM	Inner Cell Mass
iPSC	Induced Pluripotent Stem Cell
ITR	Inverted Terminal Repeat
KLF4	Kruppel Like Factor 4
LAD	Lamina Associated Domain
LIN28A	Lin-28 Homolog A
MBD	Methyl-CpG Binding Domain Protein
MEF	Mouse Embryonic Fibroblast
MET	Mesenchymal To Epithelial Transition
miPSC	Mouse iPSC
miRNA	Micro RNA

MMLV	Moloney Murine Leukemia Virus
mRNA	Messenger RNA
MYC	V-Myc Avian Myelocytomatosis Viral Oncogene Homolog
NHEJ	Non-Homologous End Joining
NSC	Neural Stem Cell
NT-ESC	Nuclear Transfer Embryonic Stem Cell
NuRD	Nucleosome Remodelling and Deacetylation
OCT4	Octamer-Binding Protein 4
OSK	OCT4, SOX2 and KLF4
ORF	Open Reading Frame
p-TEFb	Positive Transcription Elongation Factor
PAM	Protospacer Adjacent Motif
PFA	Paraformaldehyde
Pol II	RNA Polymerase II
POU	PIT/OCT/UNC Family Transcription Factors
POU5F1 (OCT4)	POU Class 5 Homeobox
PRC	Polycomb Repressive Complex
PSC	Pluripotent Stem Cell
rAAV	Recombinant Adeno-associated Virus
RNA	Ribonucleic Acid
RNAi	RNA Interference
RNP	Ribonucleoprotein
SCNT	Somatic Cell Nuclear Transfer
sgRNA	Single Guide RNA
siRNA	Small Interfering RNA
SOX2	SRY-Box 2
TAD	Topologically Associating Domain
TALEN	Transcription Activator-Like Effector Nuclease
TGFβ	Transforming Growth Factor Beta
TE	Transposable Element
TET	Ten Eleven Translocation
TMP	Trimethoprim
tracrRNA	Trans-activating crRNA
TSS	Transcription Start Site
UTR	Untranslated Region
ZFN	Zinc-Finger Nuclease

5. Introduction

Induction of pluripotency and CRISPR/Cas9-mediated genome engineering have been two of the most impactful biomedical innovations affecting the stem cell field in the past decade.

The pioneering work on cellular reprogramming by Takahashi and Yamanaka in 2006 [1], spawned a subsequent boom in human pluripotent stem cell research. As induction of pluripotency allows generation of embryonic stem cell like pluripotent cells from adult donors, this method has enabled new kinds of *in vitro* research avenues using patient derived induced pluripotent stem cells (iPSC) for disease modelling. These cells can further be differentiated into affected somatic cell types, which may not otherwise be available for *in vitro* studies, e.g., neurons or cardiac cells. Additionally, iPSCs hold great promise for an easily available source of immune matched cells for cell replacement therapies. The first human trials using iPSC derived retinal pigment epithelial cells are currently ongoing for treatment of macular degeneration [2], and other trials are to be expected. One of the strongest examples speaking for the impact of iPSC induction on medical research and the expectations set for the method has been the 2012 Nobel Prize in Physiology and Medicine awarded to John Gurdon and Shinya Yamanaka for their contributions to the cellular reprogramming field. Even though iPSCs have a huge potential for medical research and cell therapies, the derivation of the reprogrammed cells is not without issues. The method has been shown to lead to improperly reprogrammed cells and to be associated with genetic aberrations, which may affect the downstream applicability of the resulting cells. Therefore, better understanding of the processes controlling cellular reprogramming and better methods for iPSC derivation are required for faithful conversion of somatic cell types to pluripotency and recapitulation of the pluripotent cell phenotype.

The bacterial adaptive immune defence derived CRISPR/Cas9 system was first demonstrated to function as an RNA guided endonuclease in 2012 [3,4]. Although genome engineering approaches based on meganucleases, zinc-finger nucleases (ZFN), and transcription activator-like effector nucleases (TALEN) had previously been developed, the RNA guided target site determination of CRISPR/Cas9 simplified genome editing procedures. Later, CRISPR/Cas9 systems have been adapted for transcriptional control and epigenome editing. Combined with the high multiplexing capacity of the short guide RNA molecules, this makes the system very appealing for various cell biological applications, including cellular reprogramming.

This thesis work aims at developing novel cellular reprogramming approaches for improving the outcome of pluripotent reprogramming. Specifically, this includes Adeno Associated Virus (AAV)-mediated reprogramming factor delivery (**I**), CRISPR/Cas9-mediated activation of endogenous pluripotency factors (**III** and **IV**) and characterization of factors limiting reprogramming efficiency of senescent cells (**II**).

6. Review of the Literature

This part of the thesis will first describe current issues of pluripotent stem cell research followed by concepts in transcriptional control of gene expression. Thereafter, the tools and methods for cellular reprogramming and artificial control of transcription with CRISPR/Cas9 tools will be described.

6.1. Pluripotent Stem Cells

Stem cells are capable of unlimited self-renewal and differentiation into other cell types and tissues. Stem cells can be divided into embryonic stem cells and adult stem cells. Adult stem cells are multipotent and limited in their differentiation capacity and generally only contribute to the formation of a specific tissue. Adult stem cells in the body are important for the maintenance of tissue homeostasis, such as intestinal stem cells or hematopoietic stem cells, which maintain the renewal of the respective tissues. In addition to somatic stem cells, adult tissues contain germ line stem cells that produce gametes and can reconstitute the pluripotent state after conception. Pluripotent stem cells (PSC) can differentiate into all tissues of the organism. They do not exist in the adult body but can clearly be identified as a transient cell population present in early embryos. Pluripotent stem cells can be captured *in vitro* in specific culture conditions, allowing for artificial long-term maintenance of the pluripotent cell population. *In vitro* cultured pluripotent stem cells are commonly referred to as embryonic stem cells.

Due to their capacity to form all the different tissues of the body, pluripotent stem cells hold great potential for various medical applications. As PSCs can be differentiated into any tissues of the body, they are a great source for cells that cannot be easily obtained from adult tissues, such as neurons or cardiac cells. These cells can then be utilized for various applications, like drug screening, disease modelling and cell replacement therapies.

6.1.1. Derivation and Types of Pluripotent Stem Cells

Pluripotent stem cells can be considered as the *in vitro* equivalent of the blastocyst stage embryo cell populations that give rise to the embryonic tissues, i.e., epiblast precursor cells of the inner cell mass (ICM) and post implantation epiblast cells. Pluripotent stem cells can be derived either directly from early embryos as embryonic stem cells (ESC) [5–7], from germ cells [8,9], or by reprogramming somatic cells, either by nuclear transfer to oocytes (NT-ESC) [10,11], or by overexpression of specific sets of pluripotent state transcription factors (iPSC) [1]. As the derivation of ESCs generally requires the destruction of a viable embryo, human embryonic stem cell research has historically been ethically controversial. The ability to produce pluripotent cells from adult cell sources by reprogramming has helped in this regard.

Pluripotent stem cells can exist in various states of pluripotency roughly equating to different developmental time points of the embryo. Two of the most commonly studied states are referred to as primed and naïve states. These are distinguished by various parameters like morphology, gene expression, epigenetic state and ability to contribute to chimera formation [12]. Naïve state cells resemble the inner cell mass epiblast precursor cell population, whereas primed cells resemble more the single cell layer epiblast cell population of an embryo. It has been speculated that primed pluripotent cells could be more biased in their differentiation capacity towards some lineages due to intrinsic variation in priming of differentiation associated genes. In addition to the conventional naïve and primed states, a number of other alternative types of pluripotent states have been reported. These include for example the region-selective epiblast stem cells (EpiSC), which can only graft into posterior parts of post-implantation epiblast [13], various forms of 2-cell embryo-like extended potential pluripotent stem cells (EPS), which are capable of contributing to both embryonic and extraembryonic tissue development *in vivo* [14–17], and artificial pluripotent states maintained by transgene expression, like the F-class cells [18].

6.1.2. Pluripotent Stem Cell Maintenance

The maintenance of stable pluripotent stem cells in culture requires promotion of self-renewal of the pluripotent state and prevention of the differentiation of the cells. *In vitro* pluripotency is a transient state that is resolved by the cells differentiating into other tissues. The culture of pluripotent cells *in vitro* tries to capture the transient pluripotent state and the specific culture conditions are dependent on the desired state of potency. Mouse pluripotent stem cells are maintained in their conventional culture conditions in a naïve state. The naïve sate in mouse PCSs is generally maintained by LIF-signalling via STAT3 [19], and can be further promoted by culture in the presence of MAPK pathway inhibitors and Wnt activators, to inhibit differentiation of the cells and to promote active cell growth [20]. Mouse PSCs can also be derived in primed state as epiblast stem cells [21,22]. EpiSC state is dependent on FGF- and TGFβ-signalling for its maintenance. Human pluripotent stem cells resemble more the epiblast stage of embryonic cells and are also dependent on FGF- and TGFβ-signalling [23–25].

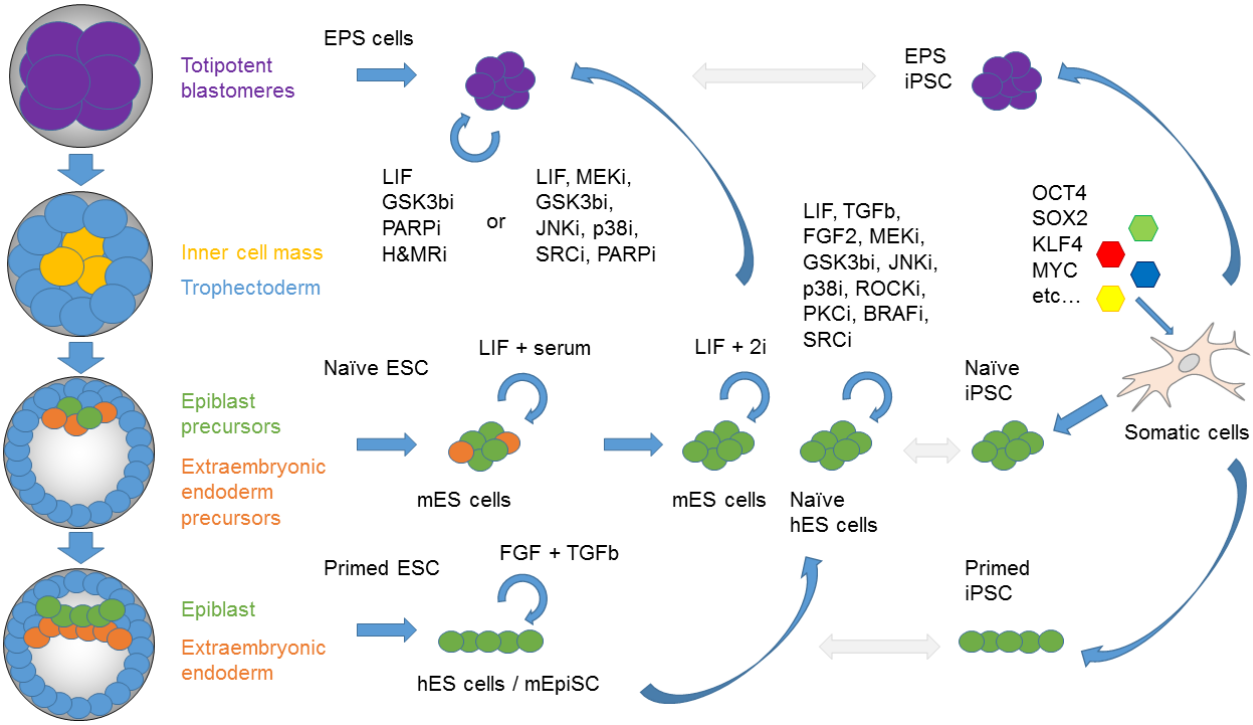

Figure 1. *Sources and maintenance of pluripotent stem cells. Embryonic stem cells can be derived from early embryos and induced pluripotent stem cells can be reprogrammed from somatic cells. PSCs cultured in different conditions roughly equate to different developmental stages of an embryo.*

FGF controls both the differentiation of naïve state cells towards primed state and the maintenance of the primed pluripotent cells. The role of FGF-signalling in the maintenance of primed pluripotent state has been attributed to its role in promoting the activation of the PI3K pathway in the presence of SMADs and the absence of Wnt signalling [26]. The maintenance of human PSCs with other PI3K pathway activating factors, such as IGF-1, Heregulin or Elabela, seems to support this notion [27,28]. FGF signalling also activates MAPK signalling and ERK phosphorylation, which is a key factor in contributing to the differentiation of naïve state cells. In primed human PSCs ERK2 has been shown to target NFY and ELK1 bound active genes to promote self-renewal and pluripotency [29]. Mechanistically ERK activation has been reported to promote Pol II Ser-5 phosphorylation at developmental genes [30]. Therefore, ERK signalling may contribute to priming of developmental genes for activation by additional signalling cues by promoting transcriptional initiation and proximal pausing of Pol II at the promoters [31]. Although it has been reported that paused Pol II is detected primarily at signalling pathway genes and not at developmental genes in mouse PSCs [32], the study used cells cultured in the presence of MEK inhibitor, which may result in loss of Ser-5 phosphorylated Pol II at these genes, if they are differently regulated by ERK, as has been suggested [30]. Combination of Hippo signalling, TGFβ-mediated promotion of Pol II Ser-7 phosphorylation and Wnt-mediated looping of enhancer and promoter regions of mesenchymal genes has been described as models for triggering mesenchymal differentiation of human PSCs [33]. The active transcription of mesenchymal genes is promoted by removal of

YAP-mediated repression of switch enhancers, that are poised, switch-like regulatory elements that allow cells to differentially interpret the same SMAD2/3 signalling [34]. As the same signalling factors, i.e. Wnt components, are required for the maintenance of naïve state cells and the initiation of primed cell differentiation, the primed cell state is not easily converted back to naïve state due to preference to differentiate.

Primed mouse EpiSCs can, however, be converted into naïve state with low efficiency by culturing in naïve state promoting culture conditions [35]. The conversion efficiency can be improved by overexpression of pluripotent factors that promote naïve state, such as Klf4, Tfcp2l1, Nanog, Esrrb, Myc, Stat3, Nr5a1 and 2, Klf2 and Prdm14 [36–43]. The conventional mouse naïve state culture conditions, containing LIF, MEK inhibitor and GSK3β inhibitor, do not support maintenance of naïve human stem cells, and only one human ESC line has reportedly been derived in these conditions [44]. However, the overexpression of a number of pluripotent state factors can help induce and stabilize naïve-like stem cell state of human cells. The human naïve conversion promoting factors contain many of the same genes as mouse, including *OCT4, SOX2, KLF4, MYC, NANOG, KLF2, NR5A2* and *RARG* [45–49]. Additionally, overexpression of YAP has been reported to promote human naïve state acquisition in part by suppressing differentiation promoting effects of GSK3β inhibition [50]. The maintenance of the human naïve pluripotent state is more demanding than mouse cell maintenance, and the culture medium needs to be supplemented with extra signalling inhibitors or activators. The methods described so far are quite variable and result in cell populations with varying level of resemblance to mouse naïve cells and human preimplantation embryo epiblast cells. Inclusion of extra small molecular compounds in the culture mix can help in converting and maintaining the human naïve state in the absence of transgenic pluripotency factor expression. The small molecular compounds and growth factors that have been used for naïve state maintenance of human cells include LIF, MEK inhibition, GSK3β inhibition, FGF signalling activation and inhibition, TGFβ signalling activation and inhibition, cAMP signalling activation, aPKC inhibition, p38 inhibition, JNK inhibition, BMP inhibition, ROCK inhibition, BRAF inhibition, SRC inhibition, Tankyrase inhibition and YAP activation [46,47,49–57]. Due to the many versions of the human naïve culture conditions, it is not yet clear what exact conditions would best support the naïve state of human cells. A recent comparison of some of the most commonly used conditions suggests that the LIF, MEK inhibitor, aPKC inhibitor and titrated GSK3β inhibitor containing conditions may be the most optimal so far, as these conditions helped maintain the cells with best resemblance to embryo epiblast cells while exhibiting the least karyotypical abnormalities [58]. However, these conditions may also still cause karyotypical abnormalities, which may be attributed to the MEK inhibitor used in the protocol, and substituting the MEK inhibitor with SRC inhibitor may help in maintaining normal karyotype and DNA imprinting of the cells in prolonged culture [59,60]. Different types of pluripotent stem cells and their maintenance is summarised in **Figure 1**.

6.1.3. Transcriptional Control of Cell Identity

Gene expression is initiated at promoter sites which define the transcriptional start sites for mRNA production. As the transcriptional profile of a cell defines the cell type, transcriptional control lies in the heart of cell identity. Cell type specificity of promoters is commonly controlled by *cis*-acting DNA elements, known as enhancers [61]. Enhancers can bind *trans*-acting factors, i.e. transcription factors, which can stimulate increased transcription from promoters, or mediate repressive function. Cell type specific expression of the enhancer binding transcription factors can therefore mediate cell type specificity of the associated promoters. The accessibility of enhancers and promoters can further be controlled by epigenetic chromatin and DNA modifications in these loci. Transcriptional activation of promoters is mediated by recruitment of the mediator complex by the enhancer bound transcription factors [62]. The mediator complex further recruits the RNA polymerase II (Pol II), which can initiate transcription of the target gene into mRNA. Enhancer recruited transcription factors therefore define cell type specific transcription.

More recently, clusters of active enhancers, deemed super-enhancers, have been described in controlling cellular identity [63]. These elements show cell type specific activity and can accommodate binding of unusually high densities of interaction partners. The super-enhancers can be established by cell type specifying master transcription factors in combination with mediator complex [64]. In pluripotent stem cells these enhancer domains are formed by the core pluripotency factors OCT4, SOX2, and NANOG.

The function of the super-enhancers has recently been proposed to follow a phase separation model [65]. This model presents separation of active enhancers and transcription promoting factors into non-membranous nuclear compartments by high density of intermolecular interactions of *trans*-acting factors and the super-enhancer loci. These types of discrete transcriptionally active compartmentalized nuclear sites with enrichment of Pol II and transcription promoting components have also been deemed transcription factories [66–68]. Similar discrete compartmentalization of nuclear processes is also known for other nuclear organelles, including nucleoli, Cajal bodies and nuclear speckles. Phase separation has also been described in driving heterochromatin formation [69], therefore it may be an integral mechanism in controlling nuclear architecture and function. Important part of the phase separation model of enhancer function is its ability to explain the vulnerability of the active super-enhancer complexes to perturbations by interference with the 'cross-linking' interactions of the *trans*-acting factors. This kind of mechanism could explain the effect that external cues have in controlling enhancer activity. Sensitivity to external factors, which mediate commissioning and decommissioning of enhancers, may therefore be the basis for controlling cell type specific expression by signalling cues, i.e., this model may explain the mechanisms by which stem cell differentiation is controlled by signalling from the cell niche.

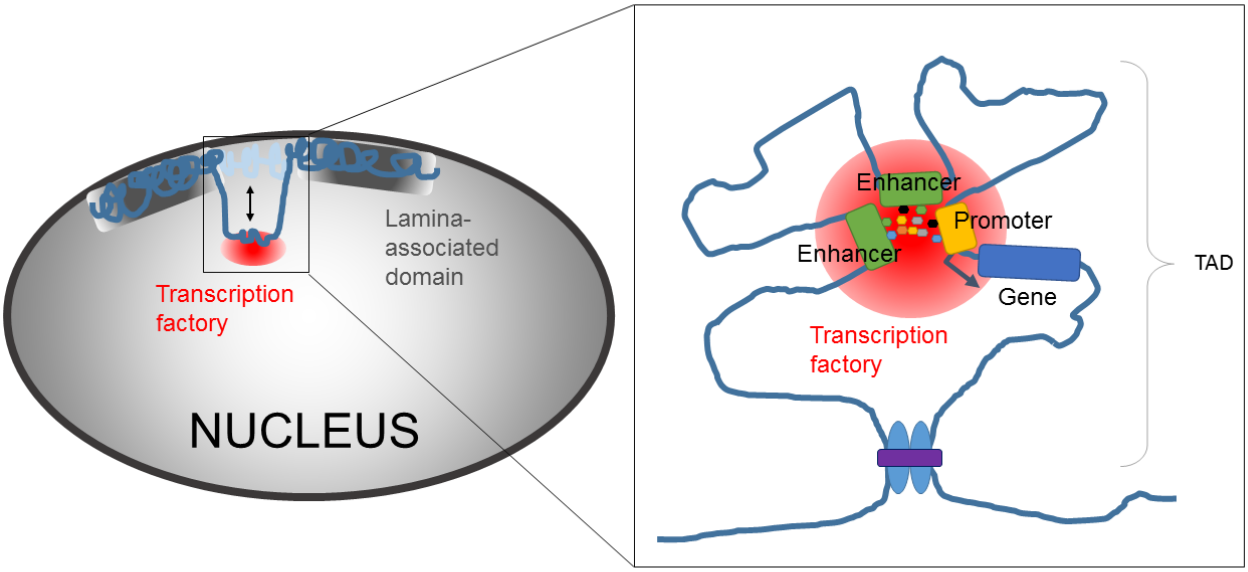

Figure 2. *Nuclear compartmentalization in transcriptional control. Repressive heterochromatin is commonly located in lamina associated domains near nuclear periphery. Upon activation gene loci can be translocated to transcription factories containing transcriptional machinery components. Transcriptional activation can be maintained by phase separation mediated by transcription factors, co-activators and RNA polymerase interacting with regulatory DNA elements. DNA is organised into topologically associating domain (TAD) loops where loop anchor contains CTCF sites and intra loop interactions are mediated by YY1 and ZNF143.*

Chromosomes in nucleus primarily occupy spatially restricted areas called chromosome territories. These are further sub-divided into smaller DNA structures with high level of intradomain interactions called topologically associating domains (TAD). The TAD boundaries are generally conserved between cell types and are marked by CTCF binding sites and cohesin occupancy. These regions, termed insulated neighbourhoods, thus appear to form a DNA sequence specific structural feature for DNA looping [70]. Gene expression control primarily occurs within these loop structures between enhancer and promoter regions located inside same TAD, intra TAD interactions forming approximately 90% of gene regulatory interactions. These DNA loop structures are generally associated with distinct patterns of histone modifications, and loop anchor sites typically bind CTCF and occur at TAD boundaries [71]. TADs are thought to form by DNA loop extrusion and boundary formation by parallel CTCF-CTCF binding motifs [72]. Possible factors mediating this process may be condensin, which has intrinsic ATP-dependent mechanochemical motor activity [73], and transcription-mediated cohesin positioning in DNA [74]. Transcription mediated TAD formation appears to be supported by presence of CTCF cohesin loops primarily in sites with higher transcription levels. However, single cell analysis of 3D chromatin structures in mouse ESCs indicated that TAD formation does not occur in all cases [75]. Physical

proximity of genomic locations in nucleus governs the interactions of enhancer and promoter sites [76]. Dynamic intradomain looping of DNA between enhancer and promoter regions can be mediated by common transcription factors such as YY1 and ZNF143 [77–79]. Intradomain promoter-enhancer interactions in ESCs have also been shown to be mediated by mediator and cohesin co-localizing with OCT4, SOX2 and NANOG, indicating a role for cell type specific transcription factors in controlling chromatin architecture [80,81]. Indeed, functional interaction of ZNF143 and OCT4 has been demonstrated in controlling *NANOG* promoter activity [82]. On the other hand, KLF4 has been shown to control long-range interchromosomal interactions in the *Oct4* locus [83]. Therefore, transcription appears to be controlled by cell type associated factors within the TADs, and longer-range interactions may further affect nuclear compartmentalization. Overall, it has been proposed that chromatin regulation can be divided into two main types of control over chromatin architecture [84]. On a larger level long-range chromatin interactions appear to be defined by open/closed characteristics of chromatin, both inter- and intra-chromosomally. On a shorter range genomic regions are more likely to interact if they share similar regulatory proteins, such as pluripotency factors or PRC complexes [84]. Therefore, a simplified view of the mechanism suggests that proteins preferably controlling short range interactions within TADs may be more important in mediating transcriptional control of target genes by mediator and cohesin complexes, and the factors preferably mediating long-range interactions may be more important in localizing genomic areas between open and closed chromatin compartments. This may further link the chromosome architecture to the phase separation model of transcriptional control.

Closed heterochromatin in cells is usually spatially segregated into nuclear lamina associated domains (LAD), near nucleoli or in pericentromeric heterochromatin [85]. LADs in cells are generally divided into cell type invariant constitutive LADs, and variant facultative LADs, which interact with nuclear lamina in a cell type dependent manner. Physical interaction of genomic loci between LADs have been shown to change in stem cells upon differentiation, reflecting the transcriptional changes of the associated genes. For example, the physical localization of multiple pluripotency genes, including *Oct4*, *Rex1* and *Nanog*, changes closer to nuclear periphery when differentiating PSCs into neural cells [86]. Additionally, targeting of genes with synthetic transactivator domains will relocate the genomic loci from LADs towards the nuclear interior in a chromatin remodelling dependent way [87]. This likely allows the loci to relocate to transcription factories upon gene activation. Knock down of YY1 can shift LADs towards nuclear interior, indicating a possible role for the factor also in heterochromatin maintenance in addition to mediating enhancer looping [77,85,88]. Nuclear compartmentalization in transcriptional control is summarised in **Figure 2**.

TAD compartments disappear in mitotic cells [75]. Therefore, chromosomal 3D structure has to be re-established at G1 phase. In order to know which genes to start transcribing after division, a set of transcription factors stay attached to the mitotic DNA. This is known as mitotic bookmarking. Many pluripotent stem cell specific factors, including ESRRB, KLF4, OCT4 and SOX2, as well as epigenetic

modifications have been shown to have mitotic bookmarking function [89,90].Therefore, these factors may also be important in re-establishing the pluripotent stem cell transcriptional program after cell division. Conversely, inhibition of the factors binding these epigenetic modifications, implied in the mitotic bookmarking, can be used to promote transcriptomic changes aimed at converting cell types. This kind of mitotic de-bookmarking has been used for example to mediate induction of pluripotency by inhibiting fibroblast transcriptional program [91]. It has also been argued that mitotic exclusion of transcription factors may be a cell fixation artefact and that many transcription factors will bind mitotic DNA in a dynamic manner [92]. In addition mitotic cells appear to maintain their transcriptional program, albeit at lower levels [93]. Therefore it is possible that cell type specific transcription in mitotic and interphase cells is primarily different by the magnitude of transcription, which may be due to more stable transcription factor interphase interactions with DNA mediated by transactivation domain interactions [92].

6.1.4. Transcription Factors Regulating Pluripotency

There are two major models for the maintenance of the mammalian pluripotent state. These models depict the pluripotent state either as a ground state or a balanced state. In the ground state model the pluripotent state is seen as an intrinsically stable state in which the self-regulatory nature of the pluripotent factors can maintain the expression of the pluripotent gene network in the absence of external signalling cues [12,20]. This interpretation is more prominent in the context of mouse pluripotent stem cells, which can be more readily maintained in a naïve state of pluripotency. This has been predicted to be linked to the transcriptional control of diapause in rodents. The other model describes pluripotent state as a balanced state between opposing pro-differentiation cues [94]. This kind of 'see-saw' model is based on the fact that overexpression of pluripotency factors can promote destabilization of pluripotent state and differentiation of pluripotent cells. For example, ectopic expression of SOX2 or NANOG can promote differentiation of ESCs into neural and mesendodermal lineages respectively [95,96]. In the balance model the opposing effects of divergent lineage specifying transcription factors form a cross-regulative network in which the balancing of antagonistic pro-differentiation cues results in the emergence of pluripotent differentiation capacity. This model is possibly more suited for the primed pluripotent state, predominant in human PSC culture, in which cells are poising the differentiation associated genes for activation. It may thus be that the differentiative capacity of the pluripotency factors results from excessive activation of poised genes in the presence of excessive amounts of particular pluripotency factors.

The gene regulatory network for pluripotency in human embryonic stem cells was initially described to be governed by the core transcription factors OCT4, SOX2 and NANOG [97]. These factors are also required for proper embryo development. OCT4, SOX2 and NANOG were found to co-occupy many of their target sites, with over 90% of OCT4 and SOX2 bound promoter regions being also bound by NANOG. These factors occupy

sites in multiple genes which have been implicated in embryo development, including signalling and transcription factors controlling differentiation and pluripotency, as well as OCT4, SOX2 and NANOG themselves. Therefore, the core pluripotency factors form a basis for a regulatory network governing the expression of differentiation associated factors as well as self-regulatory loops maintaining their own expression and pluripotency. OCT4 and SOX2 bind DNA as a heterodimer in closely spaced composite motifs containing POU domain (OCT4) and HMG domain (SOX2) recognition sites [98]. The interaction between HMG domains and POU domains has been evolutionally conserved [99]. One way of controlling the differentiation of pluripotent cell is a switch between SOX2 and SOX17 expression, which redistributes OCT4 binding to distinct 'compressed' composite DNA motifs in endodermal differentiation [100,101].

In addition to OCT4, SOX2 has been shown to interact physically with NANOG [102]. Interaction with the core pluripotency regulatory factors and co-occupancy of genomic sites has been used to identify additional factors contributing to the pluripotency maintenance network, like SALL4, ESRRB, TET1, TET2 and PRDM14 [103–107]. Further integration of transcriptional and binding data of transcription factors has expanded the pluripotency regulatory network to include factors outside of the core set, like SMAD1, STAT3, TCF7, NR5A2 (LRH1), NR0B1 (DAX1), NACC1, ZFP42 (REX1), ZFP281, TBX3, TCF7L1 (TCF3), KDM3A and TRIM24 [108–111]. More comprehensive analysis of gene expression differences in numerous tissue types, larger set of transcription factor binding data and transcriptomic perturbations of transcription factors has led to the characterization of even wider set of candidate tissue type defining master regulatory factors, some of which may contribute to the pluripotent gene regulatory circuitry as well as cellular reprogramming [112–115]. Notable pluripotency factors in these sets of candidates include factors like ZSCAN10, OTX2, ZIC2, ZIC3, MYCN, FOXH1, NR6A1, LIN28A, LIN28B, FOXO1, RARG, MYB, RORA and SOX21. Overall, these transcription factors form a widely interconnected regulatory network, which maintains the pluripotent state. The network still needs to be responsive to signalling cues to mediated context dependent differentiation of pluripotent cells. Additional differences in factors contributing to the pluripotent gene regulatory network may result from species specific differences in pluripotency maintenance and the type of pluripotent state, i.e., primed or naïve state.

6.1.5. Epigenetics of Pluripotent Cells

Epigenetics is a layer of control over DNA, which defines how DNA is interpreted without affecting the genomic sequence. Epigenetics is generally mediated by chemical modifications of DNA or by controlling the composition, positioning and tail modifications of histones, around which DNA is wrapped [116]. On a general scale epigenetics functions to divide chromatin into active euchromatin and repressive heterochromatin. Epigenetic modifications of DNA and histones are written by enzymes that catalyse the addition and removal of chemical groups and read by proteins that recognise these modifications.

The primary chemical modifications in mammalian DNA are methyl groups and their subsequent derivatives on cytosine residues [117], but other rare modifications, like adenine methylation, have also been reported [118]. Cytosine methylation is catalysed either by the maintenance DNA methyltransferase DNMT1 or by the *de novo* DNMT3 DNA methyltransferases. DNA demethylation can be mediated either by passive demethylation by cell replication or by active DNA demethylation by DNA damage repair following 5-methylcytosine (5mC) oxidation by TET enzymes into 5-hydroxymethyl- (5hmC), 5-carboxyl- and 5-formylcytosine [117]. The effect of DNA methylation on gene expression is mediated either by directly affecting the DNA binding affinity of transcription factors to the methylated sequences, or by recruiting methyl-CpG binding factors. These factors, like MeCP2 and MBDs, can further recruit other repressive complexes, like the SIN3A histone deacetylation (HDAC) complex or histone-lysine N-methyltransferase SUV39H1, that catalyses histone 3 lysine 9 trimethylation, to the methylated DNA sites [117,119].

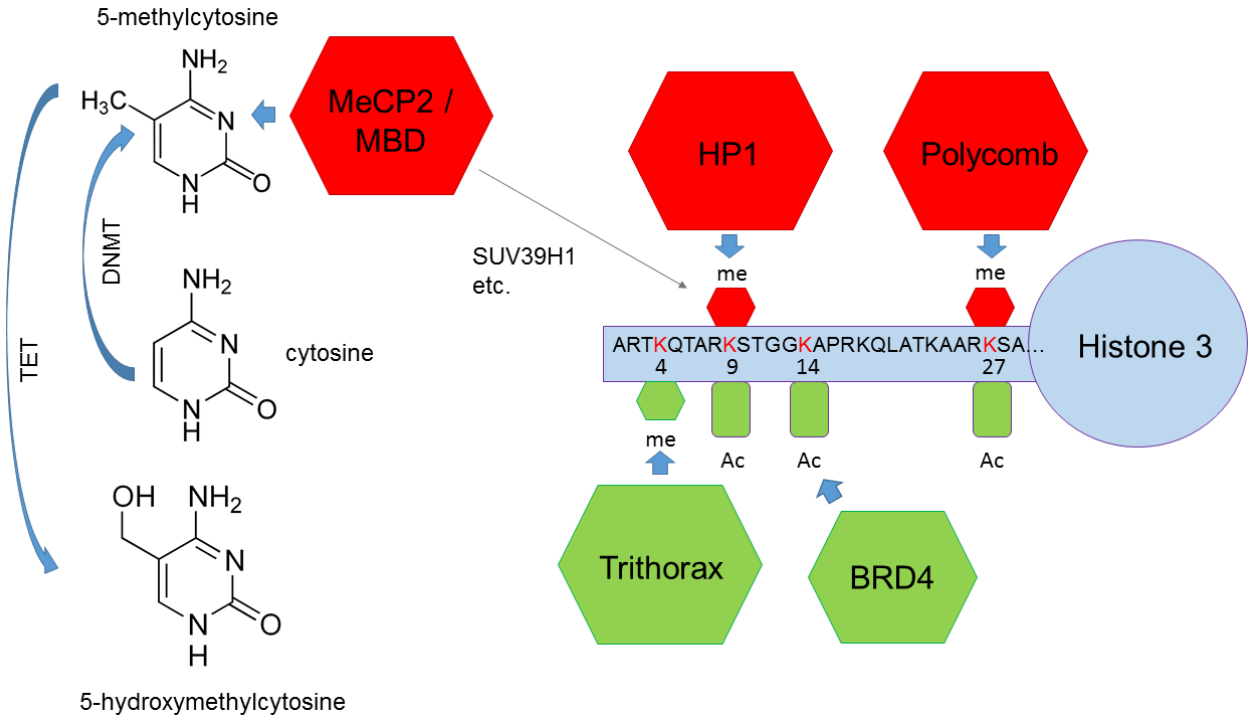

Figure 3. *Major epigenetic modification in pluripotent cells. DNA modifications are primarily targeted to cytosine residues and are catalysed by DNMT enzymes. DNA demethylation can occur via hydroxylated intermediate catalysed by TET enzymes. Methylated cytosine residues can recruit histone methyltransferases that promote repressive heterochromatin formation by HP1. Histone 3 methylation at lysine 4 and lysine 27 control promoter activity via Trithorax and Polycomb complexes respectively. BRD4 promotes active transcription.*

Histone tails can be modified in various ways including methylation, acetylation, phosphorylation, ubiquitination, sumoylation and others [120]. These modifications form a histone code that is specifically

recognised by other effector proteins. The most widely studied of these modifications are histone arginine and lysine methylation [120,121], and histone lysine acetylation [122]. Methylated and acetylated histone tails can be recognised by chromodomain and bromodomain containing factors respectively, which mediate the effect of the modified histone tails on gene expression. Notable complexes associated with histone lysine methylation are the histone 3 lysine 4 trimethylation (H3K4me3) associated Trithorax complex and the histone 3 lysine 27 trimethylation (H3K27me3) associated Polycomb comlex [123]. H3K4me3 marker is commonly associated with active gene promoter areas whereas promoter H3K27me3 is associated with gene repression. These two modifications can also occur simultaneously in a bivalent manner on a gene promoter poised for activation [124,125]. This way the cell can switch on or off a gene for example according to its differentiation path. Third notable histone modification is the methylation status of the histone 3 lysine 9 (H3K9) residue. This residue is primarily methylated by the KMT1 members G9a, SUV39H and SETDB [121], and in its trimethylated state (H3K9me3) binds the heterochromatin protein HP1 (CBX in mammals) [126], which mediates the phase separation and packaging of repressed DNA into heterochromatin [69]. Histone acetylation is commonly associated with active gene expression. On one hand, histone acetylation prevents the methylation of the corresponding lysine residue, preventing for example repressive marker addition. On the other hand, acetylated lysine residues are recognized by transcription promoting bromodomain containing factors. For example, BRD4 binds histone 3 lysine 14, histone 4 lysine 5 and histone 4 lysine 12 acetylated histones and mediates histone 3 lysine 122 acetylation, promoting nucleosome eviction from chromatin [127]. BRD4 also promotes P-TEFb activity leading to increased Pol II Ser-2 phosphorylation and Pol II pause release [128–130]. Specific genomic elements can also be defined by the histone marks they bear. Commonly used markers for defining promoter regions include the previously mentioned H3K4me3 and H3K27me3, whereas histone 3 lysine 4 mono and di methylation (H3K4me1/2) and histone 3 lysine 27 acetylation (H3K27Ac) can be used to distinguish active enhancers and histone 3 lysine 36 trimethylation (H3K36me3) is commonly found at expressed gene bodies [116,121]. Major epigenetic modifications of pluripotent cells are summarised in **Figure 3**.

Overall, pluripotent stem cells have more open chromatin structure than somatic cells [131,132]. This is characterized by large poorly defined regions of HP1 in pluripotent cells compared to discrete foci in differentiated cells and hyperdynamic plasticity of chromatin proteins [133], and widespread transcription of coding and non-coding genomic regions [134]. This open heterochromatin state is controlled in part by Nanog and Sall1 [135], and possibly by active chromatin modelling by pluripotent state specific chromatin remodeling factors [132]. The globally more permissible chromatin structure in pluripotent cells may function to make chromatin more accessible for epigenetic priming or transcriptional activation upon differentiation, but it may also require additional repressive mechanisms to prevent activation of differentiation associated genes [132].

6.2. Cellular Reprogramming

The term cellular reprogramming refers to the process of converting one type of differentiated cell into either a stem cell or into another type of differentiated cell. However, cellular reprogramming is most commonly used to refer to induction of pluripotency, and conversions between differentiated cell types are referred to as transdifferentiation. Cellular reprogramming into pluripotent state can be achieved by somatic cell nuclear transfer (SCNT), fusion with embryonic stem cells or overexpression of defined sets of transcription factors. The reprogramming process is intrinsically an epigenetic reprogramming event by which the transcriptomic program of a cell, which defines the cellular phenotype, is reset and replaced with a transcriptomic program of another cell type. This type of process was initially demonstrated by John Gurdon in the sixties by transfer of somatic cell nuclei into enucleated frog eggs. This lead to successful development of the frog embryo and demonstrated the fact the differentiation did not require changes in the genetic code of the organism [136]. Later SCNT was also demonstrated to work with mammalian cells by the cloning of the sheep Dolly [137]. The SCNT experiments indicated that the components driving the process of nuclear reprogramming in the oocytes were present in the oocyte cytoplasm. Although SCNT results in the emergence of a totipotent cells, cell fusion experiments using pluripotent cells demonstrated the presence of a pluripotent reprogramming factor set in embryonic stem cells [138,139].

The ingenious approach by Takahashi and Yamanaka aimed at defining this set of pluripotency associated reprogramming factors by combinatorial screening of a limited set of candidates [1]. Similar approaches have since been taken to find alternative sets of factors which can mediate transdifferentiation of somatic cells into other lineages, including neurons, hepatocytes and cardiac cells [140–144]. Newer more sophisticated methods rely on predicting gene regulatory networks governing cell type specific gene expression. This is done to limit the number of candidate reprogramming factors while simultaneously aiming to maximize the effect of the factors on the gene regulatory network. These types of approaches have yielded encouraging results and will likely aid in developing more efficient ways of controlling cellular phenotypes [113,145,146].

6.2.1. Induced Pluripotent Stem Cells

The specific set of factors capable of converting differentiated somatic cells to pluripotent cells was first described by Takahashi and Yamanaka in their pioneering paper in 2006 [1]. They initially screened a set of 24 selected pluripotency associated candidate factors for their ability to activate Fbx15 locus targeted Neo selection marker in mouse embryonic fibroblasts (MEF). Transduction of MEFs with a mixture of the 24 factors resulted in the emergence of G418 resistant colonies. No single factor was able to convert the cells to pluripotency by themselves, indicating that a combination of factors was needed for the process. Removal of individual factors from the set of 24 factors revealed a smaller set of 10 factors which were required for colony formation within the first 10 days of induction. Further removal of individual factors from this set of

10 factors pinpointed the factors Oct4, Sox2, Klf4 and c-Myc, which were required for successful induction of pluripotency. The combination of the four factors was also able to reprogram adult mouse tail tip fibroblasts. This core set of reprogramming factors is nowadays commonly referred to as 'Yamanaka factors', and has been subsequently used to reprogram somatic cells of various types and species. Although the initially described iPSCs showed pluripotent characteristics, many of the first iPSC clones were not fully reprogrammed. The induction of germ-line competent iPSCs were shown to be mediated by changing the Fbx15 selection into a more specific Nanog selection [147]. Finally, the most stringent measure of pluripotency of iPSCs was demonstrated by tetraploid complementation assay [148]. As tetraploid cells in an embryo can only contribute to the development of the extraembryonic tissues, the whole embryo has to be derived from the engrafted pluripotent stem cells.

Shortly after the induction of pluripotency was described in mouse, it was also replicated using human cells. This was initially demonstrated in three papers using either the same set of 'Yamanaka factors' [149,150], or a set of OCT4, SOX2, NANOG and LIN28, later deemed 'Thomson factors' [150]. The induction of pluripotency in human cells demonstrated the robustness of the reprogramming approach and conservation of the mechanisms behind the process between mouse and human. It was also revolutionary by allowing derivation of patient specific pluripotent stem cell lines for disease modelling without SCNT, as this abrogated many of the ethical hurdles regarding the use of embryos. Induction of pluripotency was also technically easier than SCNT, making it more accessible to larger research community.

6.2.2. Requirements of Reprogramming Factor Function

In general, reprogramming factors should fulfil two major criteria: they should function both as master transcription factors and as pioneering factors. Master transcription factors are factors which define a cell type specific gene expression program. These factors can determine the binding specificity of other more common transcription factors. This type of function has been shown for example for Oct4, PU.1 and MyoD in controlling Smad3 binding specificity in pluripotent cells, pro-B cells and myotubes [151]. This way, the cell type specific transcription factors can control cell type specific effects of common transcription factors, for example the cell type specificity of cellular signalling, which generally uses common downstream effectors.

Pioneering factors are defined as transcription factors which can bind into chromatinized DNA sequences, where binding sites may be occluded by histones and higher-order chromatin structures, and mediate local DNA opening. Therefore, pioneering factor occupancy at genomic loci has been proposed as a mechanism for priming genes for competency for transcriptional activation. This type of function has been described for example for FoxA and GATA factors [152], both of which are used in hepatic transdifferentiation methods [141]. Moreover, some FOX factors, like FoxA1 and FoxI1, have been shown to associate with mitotic chromatin,

indicating an additional bookmarking potential for these factors [153,154]. It is therefore likely that pioneering factor properties are required for reprogramming factors to mediate efficient targeting of genes that are to be activated in the reprogramming process. Additionally, mitotic bookmarking activity may further help maintain the desired transcriptional changes upon cell proliferation. The pluripotent reprogramming factors Oct4, Sox2 and Klf4 have also been shown to have pioneering factor activity, being able to bind to nucleosomal DNA with partial recognition motifs at the surface of nucleosomes [155]. The pioneering activity of these factors may therefore be an important part of their function in reprogramming somatic cells to pluripotency.

6.2.3. Factors Used for iPSC Derivation

The induction of pluripotency was initially described using the factors OCT4, SOX2, KLF4 and C-MYC in mouse and human cells or OCT4, SOX2, NANOG and LIN28 in human cells. The initial work was focused on characterizing the minimal set of factors which could reprogram cells, and further development has been put into replacing the reprogramming factors with other means to reduce this set of factors even more. There are still numerous different factors, which can replace the canonical 'Yamanaka' or 'Thomson' sets, as long as they meet the functional requirements for the reprogramming process. The following section will describe some of the reprogramming factors, their function in reprogramming and their substitutes in more detail.

6.2.3.1. POU5F1 (OCT4)

Oct4 is an important factor in embryo development. It is expressed maternally in the oocytes and the zygotic *Oct4* is activated at embryonic genome activation stage [156]. Thereafter Oct4 expression is restricted to the epiblast and germ cell lineages in mouse. Disruption of Oct4 expression in mouse embryos leads to failure in ICM formation and diversion of the Oct4 negative cell population to a trophoblast fate [157]. In human embryos OCT4 expression is not as clearly restricted, and OCT4 can be seen expressed in both trophectodermal and ICM compartments [158]. Accordingly, OCT4 expression seems to be more commonly required in human embryos for proper development, as knockout of OCT4 expression in human embryos leads to failure in both ICM and trophectoderm formation [159]. Curiously, even though OCT4 expression is required for proper embryo formation, maternal Oct4 is not needed for the establishment of totipotency [160]. Therefore, it is possible that the totipotent nuclear reprogramming machinery, mediating EGA, does not depend on Oct4 function, and that Oct4 is needed only for the later stages of embryo development, i.e. differentiation into trophectodermal and ICM lineages and establishment of pluripotency. OCT4 knockdown in human pluripotent stem cells leads to disruption in the maintenance of pluripotency and upregulation of trophectodermal and endodermal genes [161].

In pluripotent reprogramming OCT4 is probably the most crucial reprogramming factor [162]. Ectopic expression of just OCT4 can be used to convert neural stem cells into induced pluripotent stem cells [163]. OCT4 appears to be a unique reprogramming factor in the 'Yamanaka' set as it is the only one, which cannot be readily replaced by other OCT family members. OCT4 mediated maintenance of pluripotency and reprogramming function appears to be dependent on its heterodimerization with SOX2. A related factor OCT6 (POU3F1) cannot normally reprogram somatic cells to pluripotency. Introduction of point mutations in the OCT6 protein, which change its sequence binding preference from palindromic *OCT-OCT* motifs to heterodimeric *SOX-OCT* motifs and promote its interaction with SOX2, can convert OCT6 into a reprogramming factor, whereas introduction of the same OCT6 mimicking mutations into the OCT4 sequence abolishes the OCT4 reprogramming function [164]. The interplay between OCT4 and SOX2 thus appears to be important for successful induction of pluripotency. Conversely, limiting the availability of OCT4 to the initial phase of reprogramming can bias the cell fate towards neural lineages and be used to induce neural stem cells [165].

Some factors, such as NR5A2, TET1 or SALL4 and NANOG, which have been used to replace OCT4 in the pluripotent reprogramming factor mixture, have been implicated in the activation of the endogenous *OCT4* gene [162,166–168]. OCT4 has also been reported to be replaceable by E-cadherin over expression [169], or by cAMP signalling activation [170]. A particularly interesting group of OCT4 replacing reprogramming factors is formed by lineage specifiers. OCT4 has been reported to be replaceable by GATA3 fused with a VP16 activation domain in human reprogramming [171], and by multiple mesendodermal factors in mouse reprogramming [172]. This suggests a role for these factors in countering the neuralising effect of ectopic SOX2 expression in the absence of OCT4. It also supports interpretation of the reprogramming process as a balancing equilibrium between different differentiation pathways, similar to what has been proposed for the maintenance of the pluripotent state [94].

Pluripotency maintenance and reprogramming function of OCT4 seems to primarily be due to its association with transcriptional activation. Fusions of OCT4 with activator domains support induction of pluripotency, whereas repressive domains prevent reprogramming [173]. Additionally, versions of OCT4 fused with extra activation domains, i.e., MYOD, YAP, VP16 and MYC transactivation domains, have been described to have enhanced potential in cellular reprogramming [174–177].

6.2.3.2. SOX2

In the early embryo development Sox2 is expressed maternally at the oocyte. Subsequently zygotic Sox2 expression is seen throughout the preimplantation embryo development in various tissues. Although in mouse Sox2 expression is primarily restricted to the ICM lineage at the blastocyst stage, maternal Sox2 is important for the trophectoderm development, as Sox2 knock down in mouse embryos leads to

developmental arrest after compaction at the morula stage [178]. In blastocyst, Sox2 expression becomes primarily limited to the epiblast cell population, and primitive endoderm cells show upregulation of Sox17. After gastrulation Sox2 expression gets restricted mainly to the neuroectodermal cell populations, however Sox2 has additional roles later on in embryo development in both endodermal and mesodermal tissues [179]. At neural differentiation additional SoxB1 family members Sox1 and Sox3 become upregulated [179,180].

SOX2 can be replaced in pluripotent reprogramming with certain other members of the SOX family. These include the other neural SOXB1 members, SOX1 and SOX3, and a SOXG member SOX15, which shows reduced reprogramming capacity compared to SOX2 [181]. In addition to SOX1 and SOX3, other neuroectodermal factors, like GMNN and ZNF521, have also been used to replace SOX2 in reprogramming [171,172], similar to OCT4 replacement with mesendodermal factors. SOX17 cannot normally replace SOX2, however, introduction of point mutations to the SOX17 sequence, which mimic the SOX2 interaction with OCT4, can convert SOX17 into a pluripotent reprogramming factor [182]. This further emphasises the importance of the OTC4-SOX2 interaction in the context of pluripotency. Apart from pluripotent reprogramming, SOX2 alone can also be used for induction of neural cells from fibroblasts in permissive conditions [183].

In pluripotent reprogramming exogenous SOX2 expression appears to be temporally required throughout the early- and mid-stages of the reprogramming process [184]. This may be due to the late activation of the endogenous *SOX2* locus in fibroblast reprogramming [168], and the requirement of exogenous SOX2 protein until the processes maintaining endogenous SOX2 expression have been properly established.

6.2.3.3. KLF4

KLF4 is expressed in the early embryo, starting at the two cell stage in mouse embryos and at the eight cell stage in human embryos [185,186]. Klf4 expression in an embryo coincides with the expression of Klf5. In mouse embryos Klf2 is also expressed and restricted primarily to the epiblast cells, whereas in humans Klf2 expression is replaced by KLF17 [186]. Klf5 knock out in a mouse embryo results in defects in pre-implantation development, particularly in the ICM and epiblast compartments [187]. Klf4 knock out shows post-natal lethality, indicating a possible redundancy between the different KLF factors in the early embryo [185]. In the maintenance of *in vitro* pluripotent state, Klf2, 4 and 5 show redundant function, and a triple knock down of the factors results in impaired maintenance of the pluripotent state [188]. Later in development Klf4 is expressed in various tissues, particularly in epithelial and endothelial cells.

In reprogramming KLF4 can be replaced with KLF2 and KLF5 [181,189]. KLF4 overexpression is not essential for successful reprogramming, as it has been replaced by other factors, notably in human reprogramming with the 'Thomson' set [150]. Transgenic KLF4 replacement has also been described with various small molecular inhibitors and signaling factors, like BMP signaling [190,191]. The role of BMP signaling in KLF4 replacement is of

particular interest as BMP has been linked to mesenchymal to epithelial transition (MET) of the reprogramming cells [192], and KLF4 requirement in reprogramming has been proposed to be related to its function in promoting epithelialization of the reprogramming cells [193]. In line with the role of KLF in MET, KLF4 is also replaceable with TGFβ signaling inhibitors, which prevent TGFβ-mediated epithelial to mesenchymal transition [193]. It may thus be that the other KLF family members, capable of replacing KLF4 in reprogramming, also mediate their effect via promotion of MET. In this context the characterization of KLF17 function in human pluripotent reprogramming is of interest, as KLF17 has been implicated in epithelial phenotype in breast cancer metastasis and it is expressed at the early human embryo in place of KLF2 [186,194].

Even though exogenous KLF4 expression can be redundant in reprogramming, the expression level of KLF4 has been implicated in the progression of the reprogramming process. Reduced KLF4 expression has been described in inducing partially reprogrammed iPSCs paused at intermediate stages of reprogramming [195]. On the other hand, persistent KLF4 expression levels or improper transgene silencing may be associated with aberrant reprogramming outcomes, like defects in iPSC differentiation [196,197]. Curiously, the differentiation defective state also shows increased activation of specific LTR7/HERV-H retroviruses, which have been implicated in a naïve-like stem cell state in primates [198,199]. This phenotype may thus be somehow related to the function of KLF4 in promoting primed to naïve state conversion of pluripotent stem cells [36].

6.2.3.4. MYC (C-MYC)

MYC is a well-known oncogene that is part of a family of MYC genes consisting of MYC (C-MYC), MYCN (N-MYC) and MYCL (L-MYC). It was initially found by integration of an oncogenic virus in birds leading to the miss expression of the MYC gene. MYC is also commonly found either translocated or amplified in cancer, leading to its increased expression. MYC expression can promote cell proliferation, which is one of the mechanisms that can contribute to its oncogenic potential [200]. MYC binds to E-box motifs in DNA as a heterodimer with MAX [201]. MYC can recruit histone acetyl transferases and P-TEFb to its binding sites to mediate transcriptional activation [202], or function as a context dependent repressor by interacting with factors which may displace co-activators and recruit co-repressors [200]. As a transactivator, MYC has been shown to regulate Pol II transcriptional pause release [203]. This way MYC expression can function as a universal amplifier by promoting transcriptional output of expressed genes [204,205].

Myc is important for many developmental processes in mammalian embryo. Particularly, Myc has been shown to drive endogenous cell competition in early embryo, in which high Myc expressing cells are favoured over low expressing ones leading to elimination of low Myc expressing cells in mosaic embryos [206]. In embryonic stem cells, c-Myc and n-Myc double knockout has been shown to induce a diapause-like biosynthetically quiescent state [207]. However, this state has been described to show dissimilarities compared

to diapause epiblast cells and paused pluripotent cells induced by mTOR inhibition [208]. The transcriptional network controlled by MYC in PSCs appears to be separate from the core pluripotency network, which may be why MYC knockout does not impair pluripotency. The activity of the MYC network has been linked to cell proliferation and appears to be the cause of similarities of stem cell and cancer transcriptional networks [209].

Somatic cells can be reprogrammed in the absence of C-MYC, albeit with reduced efficiency, or C-MYC can be replaced by other members of the MYC family, N-MYC and L-MYC [181]. A lot of effort was initially put into replacing MYC in reprogramming due to its known function as an oncogene and association of increased tumour risk with incomplete MYC transgene silencing [147]. The development of non-integrative reprogramming methods has reduced the risk of using C-MYC as a reprogramming factor. In the reprogramming process MYC binds to partially degenerate E-boxes near OCT4, SOX2, and KLF4 (OSK) clusters and mediates OSK binding to nucleosomal DNA [155,210]. As OSK appear to function as pioneering factors, guiding MYC binding in reprogramming, it may be relevant to control the stoichiometric amounts of MYC compared to OSK to limit off-target binding and to promote optimal reprogramming effect of MYC. In pluripotent cells the number of OSK and MYC co-bound sites are reduced compared to early induction cells, indicating a possible change in MYC function from induction to pluripotent state [210].

6.2.3.5. NANOG

NANOG expression in the embryo occurs in the ICM and is later restricted to the epiblast cells. The knockout of NANOG in embryos causes defects both in epiblast and hypoblast development, however hypoblast defects appear to be due to lack of signalling from epiblast [211]. Although NANOG is part of the core pluripotency regulatory network it is not absolutely required for the induction or maintenance of pluripotency. NANOG negative iPSCs can be induced and maintained in the presence of vitamin C [212]. NANOG function in PSCs has been linked to pericentromeric heterochromatin regulation with SALL1 and more open chromatin in the pluripotent state [135]. It is possible that vitamin C complements for the lack of NANOG by promoting open chromatin state by KDM3A mediated H3K9 demethylation [213]. NANOG overexpression in general appears to be beneficial for reprogramming and it has been shown that NANOG can improve reprogramming kinetics by promoting acquisition of a pluripotent state in a cell cycle independent manner [214].

6.2.3.6. LIN28A (LIN28)

LIN28 is an RNA binding protein that has a role in controlling Let-7 family microRNA biogenesis [215]. LIN28 inhibits Let-7 production by binding to the terminal loop of the precursors of the Let-7 miRNA and preventing its processing into mature miRNA. Let-7 miRNAs become upregulated in cells upon differentiation, which

coincides with reduction in LIN28 expression. Let-7 miRNAs are known to target cell cycle and reprogramming associated genes such as MYC and Cyclin D1 [216]. In mouse reprogramming LIN28 has been shown to improve reprogramming kinetics in a cell cycle dependent manner [214]. This may contribute to the role of LIN28 in the 'Thomson' set of factors which lacks MYC transgene and the reported requirement of LIN28 in reprogramming of senescent cells [217]. Both LIN28A and LIN28B paralogs can be used to mediate reprogramming to pluripotency, and knocking down either of these factors reduces reprogramming efficiency [218]. In addition to the more thoroughly characterized role of LIN28 as an RNA binding protein, LIN28 has also been reported to bind DNA [219]. In this context, LIN28 is found enriched around transcription start sites, and mechanistically LIN28 appears to help recruit TET1 to genomic sites.

6.2.3.7. Other Factors

In addition to the more commonly used 'Yamanaka' and 'Thomson' sets of reprogramming factors, many other factors have been published in either replacing the commonly used factors or improving the reprogramming efficiency with the common factors. As the reprogramming factors form part of the pluripotent gene regulatory network, non-conventional reprogramming factors can aid in reprogramming by activating either the commonly used reprogramming factors or complementary parts of the gene regulatory network. Activation of reprogramming factors was demonstrated by Buganim *et al*. by replacing all of the 'Yamanaka factors' with various combinations of ESRRB, NANOG, SALL4, LIN28, DPPA2 and EZH2 [168]. Another way to replace the commonly used reprogramming factors is to use replacement factors that complement the function of the conventional factors. An example of this type of approach is the replacement of pluripotency factors by lineage specifying factors with opposing effects [172], or replacement of KLF4 with factors that promote MET [191]. Combining the common reprogramming factor activation and functional replacement approaches can also be used to replace all the common reprogramming factors. Particularly Liu *et al*. used C-JUN inhibiting protein JDP2, H3K36 demethylase JHDM1B (KDM2B), BMP downstream factor ID1 and transcription factors GLIS1, SALL4 and NR5A2 to achieve this in mouse reprogramming [220]. Other noticeable factors, which can replace more than one commonly used factors, include combinations of SALL4, SALL1, NANOG, UTF1 and MYC [221], OCT4 with BMI1 or ASF1A or TET1 [222–224], and miR302 family of microRNAs with or without mir200c [225,226].

6.2.4. Mechanisms of Cellular Reprogramming

Somatic cell reprogramming to pluripotency with defined factors is a lengthy and inefficient process compared to nuclear reprogramming by SCNT. SCNT results in complete reprogramming within one cell cycle whereas iPSC induction is a stochastic multistep process consisting of various intermediate cell populations. Consequently, completing the reprogramming process can take considerably long. The mechanisms of the

induction of pluripotency have been widely studied, which has led to better understanding of the processes governing pluripotent reprogramming. The following section will describe the main characterised phases of pluripotent reprogramming and how they form reprogramming barriers, which need to be overcome for successful reprogramming. However, these studies have primarily been made with fibroblasts using OCT4, SOX2, KLF4 and MYC transgenes, and how well these mechanisms translate to other cell types and other reprogramming factors is still a somewhat open question. The main phases of cellular reprogramming are summarised in **Figure 4**.

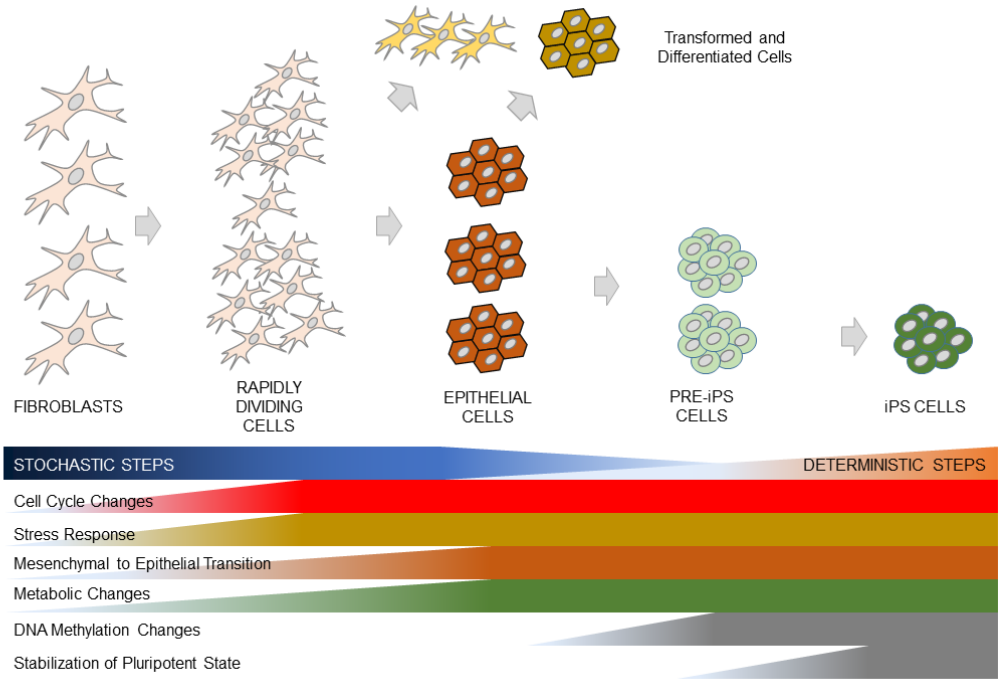

Figure 4. *Summary of the main phases of pluripotent reprogramming. Induction of pluripotency is a multistep process with stochastic events dominating the early and mid-induction phases and increase in deterministic events towards the end of the process.*

6.2.4.1. Cell Cycle Changes

The reprogramming process is initially marked by an early rapid shift in the cell proliferation rate of the reprogramming fibroblasts[227]. Active cell proliferation is generally considered to mediate the stochastic events required for proper reprogramming, and factors which can promote active cell proliferation can be used to improve reprogramming efficiency [214,216]. The role of cell proliferation in the first part of the reprogramming process is also supported by transcriptomic analysis of reprogramming cell populations [228]. The transcriptomic changes in the reprogramming cells happen primarily in two waves. The genes upregulated in the first transcriptional wave, after the first three days of induction, are mainly involved in

controlling DNA replication and cell proliferation, whereas downregulated genes are associated with cell adhesion and cell-cell contacts [228]. MYC is the likely primary factor mediating the first wave of transcriptional activation, and the rapid change in cell proliferation, by promoting activation of cell cycle promoting genes. OSK targeted genes show equal numbers of upregulated and downregulated genes in the first transcriptional wave, whereas MYC targeted genes are mostly upregulated. This may be associated with the preferential initial binding dynamics of the reprogramming factors. OSK target nucleosomal DNA at the surface of the histone octamers, mainly distally from transcription start sites (TSS), whereas MYC binds preferentially to open DNA near TSSs or along OSK in reprogramming [210]. The role of these factors in the reprogramming process may thus vary. OSK may contribute more in the initial wave to the epigenetic remodelling of their target enhancers and promoters to permissive state, as has been demonstrated for OCT4 mediated *MYOD* and *NANOG* targeting [229,230], whereas MYC ends up transcriptionally upregulating genes which already have open and accessible promoter areas. Curiously, the nucleosome depletion mediated by OCT4 has been reported to be inhibited by the DNA methylation status of the OCT4 target regions [230]. The OCT4 inaccessibility to methylated DNA may therefore partially explain the need for early cell cycle changes by proliferation mediated stochastic overcoming of reprogramming barriers, possibly by passive DNA demethylation [231]. This may also explain the effect of TET1 overexpression on reprogramming [167,224]. On the other hand, it has also been proposed that DNA replication in cell proliferation could mediate reprogramming by transiently providing DNA access for non-pioneering factors, which work co-operatively with OSK, before new nucleosomes are assembled at the newly replicated DNA [232].

6.2.4.2. De-differentiation and Mesenchymal to Epithelial Transition (MET)

The genes affected in the first transcriptional wave of the reprogramming process have been described to be KLF4 regulated, determined by the affected genes and KLF4 distribution in pluripotent cells. The KLF4 targeted genes appear to be mostly differentiation associated, and thus KLF4 has been proposed to function in the suppression of the differentiation associated genes in the first transcriptional wave [228]. However, analysis of epigenetic changes in active H3K27Ac markers in fibroblast enhancers by OCT4 and KLF4 would indicate that KLF4 targeting by itself would not lead to reduction in H3K27 acetylation, whereas OCT4 does, possibly by recruiting HDAC1 [233]. Therefore, the initial de-differentiation steps may be controlled co-operatively by both OCT4 and KLF4. This view is also supported by sequential introduction of reprogramming factors which appears to favour initial introduction of OCT4 and KLF4 [234]. Along the proposed role of the reprogramming factors in the initial de-differentiation of the reprogramming cells, co-operative OSK mediated re-distribution of somatic transcription factors away from somatic enhances, to sites elsewhere bound by OSK, has been proposed as an additional mechanism for the initial de-differentiation [233]. KLF4 has also been shown to control the epithelialization of the reprogramming cells [193]. MET is commonly considered

to be an early event in the initiation of the pluripotent reprogramming of mouse fibroblasts [192]. The transcriptional changes associated with KLF4 function in the early reprogramming may be related to its role in MET. Decommissioning the cells of their fibroblast-like transcriptional program, by inhibiting the H3K79 methyltransferase DOT1L, can also promote reprogramming by affecting the mesenchymal regulators SNAI1 and 2, ZEB1 and 2 and TGFB2 [235]. MET is not, however, an absolute requirement in the early reprogramming process, as mesenchymizing signalling factors have been reported to improve the reprogramming of epithelial cell types, and the E-cadherin / N-cadherin ratio in the starting cells seems to correlate with the reprogramming response of the cells to TFGβ signalling and TGFβ inhibition [234]. The initial role of these factors may therefore be to destabilize the starting epithelial or mesenchymal transcriptional program to help in the initial de-differentiation of the starting cells. Due to the reported roles of KLF4 in controlling long range chromosomal interactions, and the role of TGFβ in cell type-specific super enhancer formation [63,83,151], it is tempting to speculate that the molecular functions of these factors may be related to the same processes, i.e., sub-localization of their target regions into active/repressed regions in the nucleus in an epithelial or mesenchymal expression program related manner.

6.2.4.3. Stress Response

Induction of pluripotency leads to activation of tumour-suppressive mechanisms in the early stages of the process, indicating that cellular senescence is a barrier to pluripotent reprogramming [236]. As an immediate response to reprogramming factor expression, the upregulation of CDKN1A (P21) and CDKN2A (P16INK4A/P14ARF), which are potent effectors of different tumour suppressor pathways, can be seen in transcriptomic analysis of reprogramming cells [237]. The anti-proliferative, senescence-inducing and pro-apoptotic effect of the reprogramming factor expression has been widely reported [238–242]. One of the major barriers for the efficient reprogramming thus appears to be the activation of the cellular stress response, which leads to reprogramming failure. P53 controls the decision between cell death and cell cycle arrest by inducing PUMA or P21 as a response to reprogramming induced cellular stress [243]. The expression of only OCT4 and SOX2 or MYC is sufficient to induced the upregulation of p53, p21 and p19 in infected MEFs [239]. Inhibition of these pathways has been shown to alleviate the reprogramming induced senescent effect and improve the reprogramming efficiency [242]. Additionally, the reprogramming efficiency of older mouse fibroblasts can be improved by knock down of the Ink4/Arf transcript, which is upregulated in older cells [241]. Although the inhibition of these pathways can improve reprogramming efficiency, it may also risk reprogramming aberrant cells, resulting in increased aneuploidies and mutations in the resulting iPSC clones [240]. The activation of these stress pathways commonly happens as a response to DNA damage. Many possible options have been proposed for the source of the DNA damage response in cellular reprogramming, including oxidative stress, replicative stress, transcription and chromatin remodelling [244]. Different means to reduce

the reprogramming induced stress have also been proposed, including increasing the levels of CHK1 to reduce replicative stress or supplementing the reprogramming cells with antioxidants, nicotinamide or N-acetyl-cysteine [245–247]. However, the DNA damage response that inhibits reprogramming efficiency may not be completely avoidable as it might be intrinsically tied to the transcriptional and epigenetic mechanism governing reprogramming. For example, BAF60B mediated P53 activation has been reported to occur as a response to chromatin opening in transdifferentiation [248], and DNA break induced signalling has been implicated in transcriptional elongation [249]. The cellular stress response may thus be commonly activated just by forcing a drastic change in the transcriptional program of the cells.

6.2.4.4. Metabolic Switch

During the reprogramming process, cells go through a metabolic switch. This constitutes a change from primarily oxidative phosphorylation of the somatic cell to primarily glycolytic metabolism of pluripotent cells. This change in metabolism functions to accommodate the rapid requirement of ATP needed for high rate of cellular proliferation and the changes in biosynthetic needs and epigenetic processes [250]. In pluripotent cells many metabolites produced by glycolysis, including S-adenosylmethionine, NAD^+ and Acetyl-CoA, function as co-factors and substrates for different epigenomic processes. Therefore, the changes in cellular metabolism may end up directly contributing to epigenetic and transcriptional changes occurring in reprogramming. The metabolic switch in reprogramming is initiated by estrogen-related nuclear receptor-mediated burst of oxidative phosphorylation activity which leads to increase in NFR2 activity and subsequent HIF activation [251–253]. This HIF activation initiates the switch towards glycolytic metabolism. Affecting the reprogramming associated metabolic switch can also be used to promote pluripotent reprogramming by promoting the conversion into glycolytic metabolism. A good example of this is the culture of the reprogramming cells in hypoxic conditions, which facilitates the metabolic conversion and improves reprogramming efficiency [254]. Alternatively, the glycolysis promoting metabolic changes can be mediated by chemical and genetic means by PS48 and TCL1 [255,256], whereas inhibiting glycolysis can reduce reprogramming efficiency [257].

6.2.4.5. Reprogramming Phases and Heterochromatin Barriers

The reprogramming process has been described to occur in two primary phases [228]. The first transcriptional changes initially affect chromatin areas that are permissive to initial reprogramming factor binding. These preferentially targeted areas are marked by general chromatin openness [258]. The initiation of the reprogramming process appears to also be sensitive to the expression level of the reprogramming factors, as extra factors can promote the reprogramming of otherwise refractory populations and highly expressed inducible secondary reprogramming constructs can yield very high reprogramming efficiencies [228,259]. The initial stages of high efficiency reprogramming are also characterized by global reduction in repressive H3K27

methylation by increased expression of H3K27 demethylases combined with low expression of PCR2 complex members, which mediates a more open chromatin state and more plastic phenotype [259]. The reprogramming events are marked by preferential initial binding site selection by the reprogramming factors and dynamic binding site changes along the reprogramming process [233,258]. OCT4, SOX2 and KLF4 cooperate in binding site selection and the sites co-bound by OSK differ from single factor bound sites [233]. Although MYC has been reported to not affect the OSK site genomic distribution [233], it has been reported that MYC co-binding with OSK increases the OSK occupancy at OSK and MYC co-bound regions [210]. It has been proposed that the presence of MYC in OSK bound regions could help by recruiting chromatin modifying factors which may stabilize OSK binding to DNA [260]. Similar mechanism may mediate the reprogramming effect of MYC transactivation domain fusion with OCT4 [177].

After the early reprogramming events, the cells enter a mid-induction phase, which is characterized by stochastic progression of the reprogramming process. The progression of the mid phase appears to be limited by the accessibility of the reprogramming factors to the late activated genomic loci, which are enriched for pluripotency factors. A major barrier for the progression seems to be formed by DNA CpG methylation and H3K9me3 marked heterochromatin in the late pluripotency factor loci, resulting in large differentially bound regions, lacking initial reprogramming factor binding, and late activation of many CpG methylated regions [210,261]. The importance of the heterochromatin factors and histone and DNA modifications in limiting the pluripotent reprogramming efficiency is evident from the effect that targeting these factors has on the reprogramming efficiency. DNA methyltransferase inhibitors, like 5'-azacytidine, can be used to enhance reprogramming, whereas knock down of the Tet2 in MEFs abolished iPSC colony formation [237,262,263]. This indicates that DNA demethylation either by preventing re-methylation or by promoting DNA hydroxymethylation can help in overcoming reprogramming barriers. As DNA methylation primarily functions to recruit H3K9 histone methyltransferases, which mediate HP1 binding and DNA compaction into heterochromatin, DNA demethylation can alleviate gene repression. Accordingly, affecting the expression of multiple heterochromatin associated histone writers and readers can influence reprogramming efficiency. These include reduction in reprogramming efficiency by knock down of H3K9 demethylases KDM3A, KDM3B, KDM4C or increase in reprogramming efficiency by knock down of H3K9 methyltransferases SETDB1, SUV39H1, EHMT1 and EHMT2 or H3K9 binding HP1 family factor CBX3 [235,264,265], as well as increase in efficiency by over expression of H3K9 demethylase KDM4C [265]. In addition, knock down of TRIM28 (KAP1), which recruits SETDB1 to ERV elements and mediates heterochromatin formation, can promote efficient reprogramming [266]. Knock down of the replication associated chromatin assembly factor complex CAF-1 factors Chaf1a and Chaf1b can also promote mouse pluripotent reprogramming by promoting local depletion of H3K9me3 in subset of reprogramming resistant regions [267]. Curiously, these reprogramming resistant regions have initially been described in limiting the efficiency of mouse SCNT, and factors that can affect their

methylation status, like Kdm4d over expression or Suv39h1/2 siRNAs, can promote successful SCNT [268]. Therefore, the barriers limiting SNCT and pluripotent reprogramming are likely to be related.

Most powerful effects on pluripotent reprogramming can be achieved by interfering with chromatin repressive complexes. These methods can mediate high efficiency deterministic reprogramming, alleviating the reprogramming barriers so much that the stochastic phase is practically abolished and the whole cell population progresses towards pluripotency in a highly synchronous manner. This can be achieved for example by depleting the NuRD complex component MBD3 which results in nearly 100% reprogramming efficiency [269]. MBD3 has been reported to bind to hydroxymethylated DNA [270], and its function in reprogramming may be linked to the function of TET enzymes. On the other hand, the role of MBD3 as a 5hmC binding factor has been questioned, as MBD3 appears to have a higher preference for 5mC instead of 5hmC [271,272]. Therefore, the exact function of MBD3 in reprogramming is not fully clear. Curiously, MBD3 ablation has also been reported to decrease reprogramming efficiency, indicating context dependence of its function [273]. Another practically deterministic reprogramming process can be seen in B cell reprogramming with transient C/EBPα overexpression [274]. This has been linked to granulocyte/macrophage progenitor like conversion of the B cells and has been association with increased TET2 and KLF4 expression and LSD1 (KDM1A) and BRD4 function [274,275].

After the stochastic mid phase, the late part of the reprogramming process is characterized by a second main wave of transcriptional changes which are mostly affecting genes associated with embryonic development and stem cell maintenance [228]. This late phase has been deemed hierarchical so that the activation of certain late stage markers happen in a sequential manner in correctly reprogramming cells. Activation of some predictive markers, such as Esrrb and Utf1, have been proposed as potential key events in initiating the late hierarchical phase [168]. It is possible that the hierarchical phase is initiated when sufficient amount of pluripotency factors are stochastically activated to mediate more active dissolution of the reprogramming barriers and more effective targeting of late pluripotency control regions leading to full reorganization of the pluripotent regulatory network. Interestingly the proposed predictive marker Esrrb has been shown to promote rapid and more efficient co-operative binding of OCT4, SOX2, KLF4 and MYC to late accessible pluripotent enhancer loci in MEF reprogramming [233].Thus, inclusion of Esrrb in the reprogramming factor cocktail also leads to increase in reprogramming efficiency. Overall, this may suggest a possible optimal reprogramming method by which the late-stage deterministic reprogramming events can be actively promoted. This could be done by maximising the relevant set of rate-limiting pluripotent reprogramming factors to maximise the co-operative binding effect of the pluripotency factors to critical control elements. This could also be speculated to function by promoting more efficient establishment of active pluripotent cell super-enhancers and phase separation of pluripotent loci into active regions by increasing the amount of interactions in the pluripotent loci [65].

6.2.4.6. Stabilization of Pluripotency and Aberrant States

Throughout the reprogramming process the cells have been described to go through various somatic cell like states, including epidermal, primitive streak-like mesendodermal and pre-implantation-like populations [276–278]. The transient cell states and the plastic phenotype associated with mid-induction may be useful for promoting targeted differentiation after de-differentiation with the pluripotent reprogramming factors [144,279]. Towards the end of the reprogramming process, there is a transition from a pluripotent maturation state towards pluripotency stabilization. This transition is dependent on the absence of ectopic transgene expression and is marked with changes in expression of genes associated with the transient somatic- or preimplantation-like cell states [280]. Curiously, the late stabilization phase may also be associated with decrease in the chimeric capacity of the iPSCs, possibly due to more stable pluripotent state promoted by cell culture adaptation [281]. The late stabilization step is also generally associated with silencing of retrovirally delivered transgenes and independence of transgene expression for the maintenance of the pluripotent state. Persistent maintenance of the reprogramming factors has been shown to be associated with incomplete reprogramming and induction of epigenetically stable aberrant pluripotent states [18,278,280].

6.2.4.7. Other Reprogramming Barriers

In addition to the previously described reprogramming barriers, a number of other factors limiting the efficient pluripotent reprogramming of somatic cells have been reported [282]. Many of these factors are associated with ubiquitination and sumoylation processes and have been found by RNAi screens, or more recently by CRISPR screens, of reprogramming cell populations. Depletion of Small Ubiquitin-Like Modifier 2 (SUMO2) has been reported to increase both mouse and human iPSC induction efficiency [283]. Although the role of this protein in reprogramming has not been properly characterized, the authors speculate that its depletion may be related to de-repression of epigenetically silenced pluripotency loci and histone and protein synthesis genes. The depletion of the SUMO-conjugating enzyme Ubc9 (Ube2i) also promotes increase in reprogramming efficiency [267,284]. Similarly, factors functioning in protein ubiquitination have been reported to affect reprogramming efficiency, including enhanced efficiency by depletion of FBXW7, a subunit of E3 ubiquitin ligase complex [284–286], or related genes UBE2D3, UBE2E3 or RNF40 [285], and reduction in efficiency by depletion of the proteasome subunit Psmd14 [286]. This indicates that post-translational modifications play an important role in successful reprogramming and in the control of pluripotent state [287].

A number of cell adhesion and motility associated factors have been implicated in reprogramming barrier function. These include many members of the ADAM family, depletion of which results in increased human cell reprogramming efficiency [285]. A proposed function for these factors is in inhibiting an integrin switch occurring in reprogramming from fibroblasts to iPSCs. Additionally, many factors associated with endocytosis have been reported as reprogramming barriers via the TGFβ pathway [285].

Apart from the factors affecting heterochromatin formation, other epigenome editing factors have also been implicated in inhibiting efficient reprogramming. This includes the gene transcription associated H3K36 methylation marker and the respective lysine demethylases KDM2A, KDM2B [121]. Knockdown of these enzymes has a negative impact on reprogramming whereas overexpression of KDM2B promotes reprogramming with OCT4 in a PCR1.1 variant dependent manner [285,288]. The role of these factors in reprogramming may therefore be in repressing the expression of differentiation associated factors, which is supported by the OCT4 and KDM2B mediated repression of mesendodermal genes SOX17, GATA4 and TBX20 in reprogramming [288]. Similar type of function has been implicated for the transcription associated H3K79me2 and the knockdown of the respective methyltransferase DOT1L, which also represses EMT associated genes [235,289]. Interference with the active transcription associated chromatin markers may therefore contribute to more efficient loss of the mesenchymal starting state or inhibition of aberrant mid-induction mesenchymalization.

6.2.4.8. Epigenetics of Reprogramming

The reprogramming process is marked by a wide scale remodelling of the cellular epigenome. The process initiates with rapid remodelling of histone modifications in a large set of target loci followed by slower changes in DNA methylation [232,290].

6.2.4.8.1. Histones and Nucleosomes

The rapid changes in histone tail modifications are initially preferentially targeted to open and chromatinised DNA but not to heterochromatin [210]. The expression of the reprogramming factors results in rapid emergence of H3K4me2 at targeted genomic loci, a marker of poised promoters and poised and active enhancers depending on the presence of H3K27Ac [232,290]. The initial function of the factors thus appears to promote epigenetic histone modifications that can poise genes for activation or inhibition. This may be related to the role of the reprogramming factors in controlling pluripotency by priming genes to be responsive to signalling cues that mediate differentiation. Perhaps the most prominent reprogramming factor controlling this process is OCT4. Accordingly, OCT4 has been reported to interact with a wide range of active and repressive chromatin complexes. These include PRC1 complex factors, LSD1 complex factors, and various chromatin remodelling complexes including SWI/SNF family BAF complex factors like SMARCA4 and SMARCC1, CHD family NuRD complex factors like HDAC1/2 and MBD3, ISWI family factors like SMARCA5, INO80 family factors like INO80, TRRAP and EP400, and histone chaperone complex factors like ASF1A and HIRA [291–293]. Especially the roles of SMARCA4 and CHD4 have been pinpointed as important for OCT4 function in reprogramming. A single amino acid mutation in an OCT4 linker domain, which interacts with these factors, has been shown to abolish OCT4 reprogramming function [294]. The role of PARP1 interaction with OCT4 may

also be linked to its role in histone mobilisation [292,295]. It could thus be speculated that OCT4, as a pioneering pluripotency and reprogramming factor, functions to mediate chromatin remodelling of its target loci, which could promote binding of non-pioneering factors and signalling factors, while simultaneously repressing spurious activation of the sites. This could explain the effect of fusing transactivation domains with OCT4 to promote only OCT4 mediated reprogramming [174], if OCT4 by itself lacks transactivation capacity which may be promoted by other factors. It could also explain the signalling dependence of OCT4 mediated reprogramming with some of its interaction complex components, like GFD9 dependence of OCT4 and ASF1A reprogramming [223], and BMP4 interference of OCT4 and KDM2B reprogramming, which depends on PRC1.1 [288]. Additionally, the lack of sufficient amounts of signalling factors, like TGFβ signalling effectors SMAD2/3, may be a limiting step in reprogramming [296].

6.2.4.8.2. DNA Methylation

Although changes in histone modifications emerge rapidly in the reprogramming process, DNA methylation changes occur gradually. Binding sites of activated factors have been reported to exhibit focal DNA demethylation during reprogramming whereas wider neighbourhood demethylation is seen only later in pluripotent state [261]. The gradual resetting of the DNA methylation status may result in persistent epigenetic footprints of the starting somatic cell type, which are erased only in long term expansion of the iPSCs [297]. The ineffective resetting of the DNA methylation status of the somatic cell loci may cause unwanted epigenetic aberrations in the resulting iPSCs. Particularly, large scale differentially methylated regions (DMR) have been reported in human cell reprogramming, the resulting iPSC demonstrating aberrant non-CpG hypomethylation near centromeric and telomeric regions [298]. These non-CpG DMRs tend to also harbour partially CpG methylated sequences reminiscent of the starting cell type. However, these incomplete CpG modifications in the subtelomeric non-CpG DMRs appear to be caused by aberrant DNA hydroxymethylation [299], as bisulphite sequencing cannot distinguish between 5mC and 5hmC modifications [300]. Overall, the DMRs overlap with H3K9me3 markers in the starting cell populations. They also overlap with the initial reprogramming factor binding resistant regions, indicating a link between somatic cell heterochromatin status and reprogramming associated aberrations of DNA methylation [210,232,298].

Active DNA demethylation has also been implied in pluripotent reprogramming. This has been reported to be Activation Induced Cytidine Deaminase (AID)-dependent in mouse ESC and human fibroblast cell fusion heterokaryons [301]. Curiously, the role of AID in transcription factor mediated reprogramming appears to be reprogramming stage dependent. The absence of AID appears to accelerate the reprogramming in the beginning of the process, whereas AID in the later stages helps in stabilizing the pluripotent state [302]. AID functions as a cytosine and 5mC deaminase, which targets single stranded DNA and converts cytosine and 5mC to uracil and thymine respectively [303,304]. The resulting mismatches can then be repaired to mediate

cytosine demethylation. Insufficient AID activity at the later stages of reprogramming may thus limit the cell type conversion to stable pluripotent state by limiting active demethylation [302]. The early role of AID is not quite clear, but it is possible that initial DNA methylation may function in reprogramming factor targeting, as KLF2, 4 and 5 have been reported to be 5mC readers [271]. The choice of the reprogramming factors may further affect the outcome of the aberrant DNA methylation in reprogramming, as 'Yamanaka factors' may promote more CpG demethylation failures and 'Thomson factors' may promote more CpG methylation failures [305]. On the other hand iPSC CpG-DMRs have also been reported to be primarily hypomethylated and associated with both KLF4 and FOXL1 motifs [298]. Therefore, the role of the reprogramming factors in this process is still unclear. In addition to AID, active DNA demethylation can also be promoted by TET mediated 5mC and 5hmC oxidation to 5-carboxycytosine and 5-formylcytosine followed by TDG-mediated cytosine excision [306–308]. The limiting availability of TET enzymes in the beginning of the reprogramming process may therefore limit active DNA demethylation and prolong the reprogramming process.

The reprogramming process can also result in aberrations in DNA methylation in imprinted loci, in which the maternal and paternal alleles normally have different methylation patterns. The imprinting defects in human pluripotent cells were recently reported in a large scale study, which concluded that iPSCs contain more commonly imprinting defects than ESCs and that the loss of imprinting is more commonly targeting the paternal allele [309]. The reprogramming conditions may also contribute to the loss of imprinting occurring during reprogramming. For example, the imprinted mouse *Dlk1-Dio3* locus has been reported to be shielded from hypermethylation and loss of imprinting by addition of ascorbic acid, a cofactor for TET [310,311], and prolonged culture in the presence of MEK inhibition may cause aberrant hypomethylation of imprinted genes by downregulation of DNA methyltransferases [59].

6.2.4.8.3. X-chromosome Inactivation

During reprogramming of mouse cells the X-chromosome state of female cells changes from the somatic state with one inactivated X-chromosome and one active X-chromosome (XiXa) to the pluripotent state with two active X-chromosomes (XaXa). The second X-chromosome is normally silenced upon differentiation by XIST, which coats the X-chromosome with XIST RNA. This leads to subsequent compaction of the X-chromosome marked by repressive H3K27me3 foci. Unlike mouse iPSCs, that contain two active X-chromosomes, human iPSC X-chromosome activation status after reprogramming is not as definite [232]. Human iPSCs have been reported to both maintain somatic cell X-chromosome inactivation status and to activate the second X-chromosome. The differences between different reports may be caused by different culture conditions, as induction of pluripotency on LIF producing SNL feeders (STO cell line transformed with neomycin resistance and murine LIF genes) can produce human iPSC lines with XaXa status [312]. In addition, human pluripotent cells have been shown to go through loss of XIST expression, while maintaining Xi status,

and gradual partial degradation of X-chromosome silencing, deemed erosion of dosage compensation [313]. This is a process that can lead to leaky activation of X-chromosomal genes that should be silenced and can result in aberrant X-chromosome epigenetic states. Therefore, it is important to characterize the X-chromosome status of female iPSC lines, as aberrant X-chromosome inactivation may affect the behaviour of the cells in downstream applications.

6.2.4.8.4. Resetting the 3D Genome

The reprogramming process efficiently resets the global genome 3D architecture, removing the starting cell type chromatin conformation and replacing it with pluripotent cell specific conformation both in TAD and in sub-TAD level [314]. Early passage iPSCs have, however, been found to contain some cell-of-origin specific DNA topological features, which do not appear to be somatic cell derived memory, but established during the reprogramming process in a replicable cell type specific manner [314]. These atypical features do not appear to correlate with the expression of the genes associated with them, so it is possible that the topological changes can be tolerated in iPSCs as long as they do not affect the establishment or maintenance of pluripotency. The atypical domains in early passage iPSC can distinguish iPSC form different cell type sources. As the features are established in the reprogramming intermediates, this indicates a differing path for pluripotency for the different somatic cell types. Aberrantly rewired chromatin topology has also been detected in iPSC derived from NSC [315]. These interactions appear to particularly affect pluripotency genes, which are expressed at the NSC state. The affected genes appear to be hyper connected, meaning that they still retain some of the NSC specific interactions as iPSCs. These aberrant interactions also seem to correlate with CTCF sites, since extra connections can be retained in NSC-specific persistent CTCF sites while ESC-specific CTCF connections can be depleted [315]. This would suggest a model by which erroneous chromatin interactions may be retained from reprogramming intermediate states if they support expression of pluripotency-associated factors, which have been activated at the intermediate states, in the absence of proper ESC specific CTCF site associated connections. These aberrant interactions can be reset either by extended passaging or by more specific growth conditions, like 2i [314,315]. Both of these cases may be linked to DNA methylation changes, as suggested by several lines of evidence: a) naïve state culture conditions have been liked with global hypomethylation [316], b) long term expansion erases somatic cell footprints [297], and c) CTCF binding is known to be affected by DNA methylation [317].

Affecting the 3D conformation of DNA may also contribute to reprogramming efficiency. This is suggested by the observation that knockdown of YY1 improves the reprogramming efficiency of human fibroblasts [235]. It is possible that transiently downregulating the expression of YY1 may help in destabilising the enhancer-promoter interactions of the starting cell type [79], thereby promoting the initial de-differentiation of the cells. Restructuring the chromatin 3D conformation may also be important for the activation of pluripotency loci.

A recent paper studying the chromatin conformation changes in high efficiency blood cell reprogramming by C/EBPα priming described initial changes in the open vs repressed chromatin localization of genes occurring before transcriptional changes [318]. Therefore, one requirement for efficient activation of a gene, and activation of pluripotency program, may be initial reorganisation of chromatin 3D structure.

6.2.4.9. Small Molecules to Enhance Efficient Reprogramming

Small molecular compounds offer versatile tools for pluripotent reprogramming. Small molecules are easy to use as they can be supplemented in the cell culture medium, allowing good temporal control over their effects. Small molecular compounds have primarily been screened for their effect in either improving the reprogramming efficiency or replacing reprogramming factors [319]. Many of the compounds used for pluripotent reprogramming are agonists or antagonists of epigenetic enzymes or enzymes functioning in signalling pathways. Although small molecular compounds are convenient to use, they generally affect common pathways and intrinsically lack the sequence level specificity of transcription factors. Therefore, small molecules may be most useful when combined with transgenic transcription factors for affecting the reprogramming process.

Many different compounds have been studied in the context of pluripotent reprogramming. The first reports were on factors affecting the epigenetic state of the reprogramming cells, like the DNMT inhibitor (5'-azacytidine) or HDAC inhibitors (suberoylanilide hydroxamic acid, trichostatin A and valproic acid) [262]. Other molecules acting on epigenetic enzymes and improving reprogramming efficiency were later reported. These included HDAC inhibitor (sodium butyrate) [320,321], H3K9 mono and di methylase, G9A, inhibitor (BIX-01294) [322], DNMT inhibitor (RG108) [323], H3K4 demethylase, LSD1, inhibitors (parnate and RN-1) [278,324], H3K79 methyltransferase, DOT1L, inhibitor (EPZ004777) [235], and vitamin C, a cofactor for TET dependent DNA demethylation and JHDM1A/1B dependent H3K36 demethylation [325–327]. Many small molecular compounds affecting different signalling pathways have also been reported for iPSC induction, including calcium signalling agonist (BayK8644) [323], TGFβ signalling inhibitors (A-83-01 in combination with AMI-5, 616452, and SB431542 in combination with MEK inhibitor PDD0325901) [328–330], cyclic AMP analog (8-Br-cAMP) [331], Wnt signalling activator (CHIR99021) [324], SRK family kinase inhibitors [332], and retinoic acid signalling agonists (CD437 and AM580) [48]. Additionally, a set of compounds, deemed OAC for OCT4 activating compound, have been described to promote reprogramming efficiency, although the mechanism of the effect was not resolved [333]. Different combinations of the inhibitors have been used to replace many of the conventional reprogramming factors. In human reprogramming, SOX2, KLF4 and MYC can be replaced by small molecular compounds, enabling induction of pluripotency with only OCT4. This can be achieved by reprogramming epithelial human cells with OCT4 overexpression in the presence of butyrate, A-83-01, PD0325901, CHIR99021, parnate and a PDK1 activator PS48 [255]. A similar cocktail of molecules, containing valproic acid,

CHIR99021, 616452 and parnate, has been reported to promote mouse fibroblast reprogramming with transgenic OCT4 [334]. Finally, by screening for *Oct4* activating compounds, cAMP signalling activator forskolin was identified as a replacement for transgenic OCT4 [335]. By combining the chemical conditions which allow reprogramming with only OCT4 (valproic acid, CHIR99021, 616452 and parnate) with the chemical conditions which allow activation of OCT4 (forskolin), cells can be reprogrammed to pluripotency with addition of two extra molecules (S-adenosylhomocysteine hydrolase inhibitor (3-deazaneplanocin A) and a retinoic acid receptor ligand (TTNPB)) [335]. This chemical reprogramming method depends on the activation of endogenous *Sox2* and *Sall4* followed by the activation of *Gata4*, *Gata6*, *Sox17* and *Oct4* in response to the treatment with CHIR99021, 616452 and forskolin. Therefore, the pre-existing transcriptional control program in the reprogrammable somatic cells needs to be properly responsive to the chemical cues for the system to work. Related fully chemical reprogramming approaches have later been described also for various transdifferentiation approaches, primarily using mouse cells [336].

6.2.4.10. Alternative Starting Cell Types and Pathways

The majority of studies regarding the pathways to pluripotency have been made using fibroblasts. The starting cell population may affect the way the cells behave in the reprogramming process and the nature of the barriers for efficient reprogramming. Therefore, it is important to characterise the events that govern the reprogramming processes of cell types other than fibroblasts. To this end, both similarities and dissimilarities in the reprogramming processes have been reported between different starting cell types.

Analysis of the gene expression of different types of human cells has indicated various general similarities in the early, mid and late stages of reprogramming [277,337]. The early stage changes share a gradual downregulation of extracellular matrix organization associated factors and TGFβ signalling factors. Expression of innate immune response factors can be seen at early to mid-stages, which may be related to the reprogramming vector expression [337]. Genes that present a transient peak of expression in the mid-stage of reprogramming are primarily associated with cell junction assembly and differentiation-associated terms, like epidermis development. Late upregulated and gradually upregulated genes are mainly associated with neural development and cell cycle [337]. The fact that mid stage expressed genes have been associated with differentiated lineages may be due to the effect of the pluripotent reprogramming factors redistributing the somatic cell transcription factors to new sites away from their primary somatic binding sites [338,339]. The persistent expression of somatic cell specific transcription factors could thus affect the reprogramming process, particularly during the early and mid-stages of reprogramming.

The cell type of origin has also been reported to define the reprogramming path that the cells take and to affect differentially the reprogramming of different cell types [340]. Particularly genes that are downregulated or transiently upregulated are affected in a cell type specific manner. This is likely due to downregulation of

somatic cell programs in the de-differentiation phase and transient upregulation of off-target genes in the presence of persistent somatic transcription factors with the reprogramming factors. The reprogramming cells of different origins also demonstrate different order of pluripotency marker acquisition. For example, neutrophils have been reported to primarily upregulate Oct4 before becoming SSEA-1 positive, whereas keratinocytes show early expression of epithelial markers and late activation of Oct4 expression [340]. Keratinocyte reprogramming also indicates that changes in the mesenchymal and epithelial status of the cells, e.g. MET, is not required for reprogramming *per se*, but that the cells maintain their original epithelial expression program. The reprogramming of different cell types also shares certain features. For example, the main activated targets of the reprogramming factors are associated primarily with pluripotency in all cases [340]. The reprogramming culture conditions may further affect the route that cells take to become pluripotent. In the chemical reprogramming of mouse fibroblasts, the cells go through a transient extraembryonic endoderm-like state, unlike in the transgenic transcription factor mediated reprogramming [341]. Although the transient cell populations described for reprogramming intermediates resemble certain other cell types, they do not always translate to functional equivalents. For example the primitive streak-like mesendodermal state which has been described for reprogramming intermediates does not necessarily result in protein level expression of the transient state markers [277,342].

6.2.4.11. What Are the Requirements for Optimal Reprogramming?

In summary, the efficient reprogramming of somatic cells to pluripotency can be simplified to three basic principles.

(**1**) The process needs to mediate efficient decommission of the pre-existing somatic cell program. This notion is supported by the observations that the reprogramming efficiency can be increased by inhibition of factors that maintain active transcription markers in the early reprogramming, such as DOT1L and BET inhibitors [91,235], and downregulation of somatic cell specific factors, like Egr1 or AP-1 factors [339,340]. It is, however, important to keep in mind that these mechanisms maintain the proper growth of the starting cells and too low expression of these factors may decrease the viability of the reprogramming cells.

(**2**) Efficient activation of crucial facultative heterochromatin located genes. This is supported by the restricted initial binding of the reprogramming factors to some of the heterochromatin located pluripotency associated sites [210], and the fact that inhibition of repressive chromatin modifications and heterochromatin components increase reprogramming efficiency [265,321,343].

(**3**) Comprehensive targeting of pluripotency network factors. It has been reported that the targeting of the pluripotent reprogramming factors to relevant pluripotency associated on-target sites is improved by including additional pluripotency factors in the core factor set, like Esrrb [233]. Therefore, incorporation of more pluripotent gene regulatory network components in the reprogramming mixture should result in improved fidelity of reprogramming. This should also promote more efficient activation of the pluripotent state gene regulatory network by maximizing the network nodes targeted. With the recent development of CRISPR mediated gene activation, this approach should be more practically attainable (see later in chapter 6.4.).

6.2.5. Cellular Reprogramming Vectors

As the reprogramming process relies on the delivery of a defined set of reprogramming factors into the cells, many different means have been developed to achieve this goal. These include viral and non-viral delivery, integrative and non-integrative vectors and DNA, RNA and protein based transduction [344]. The following section will describe the main aspects of the various methods. A summary of cellular reprogramming vectors is presented in **Figure 5**.

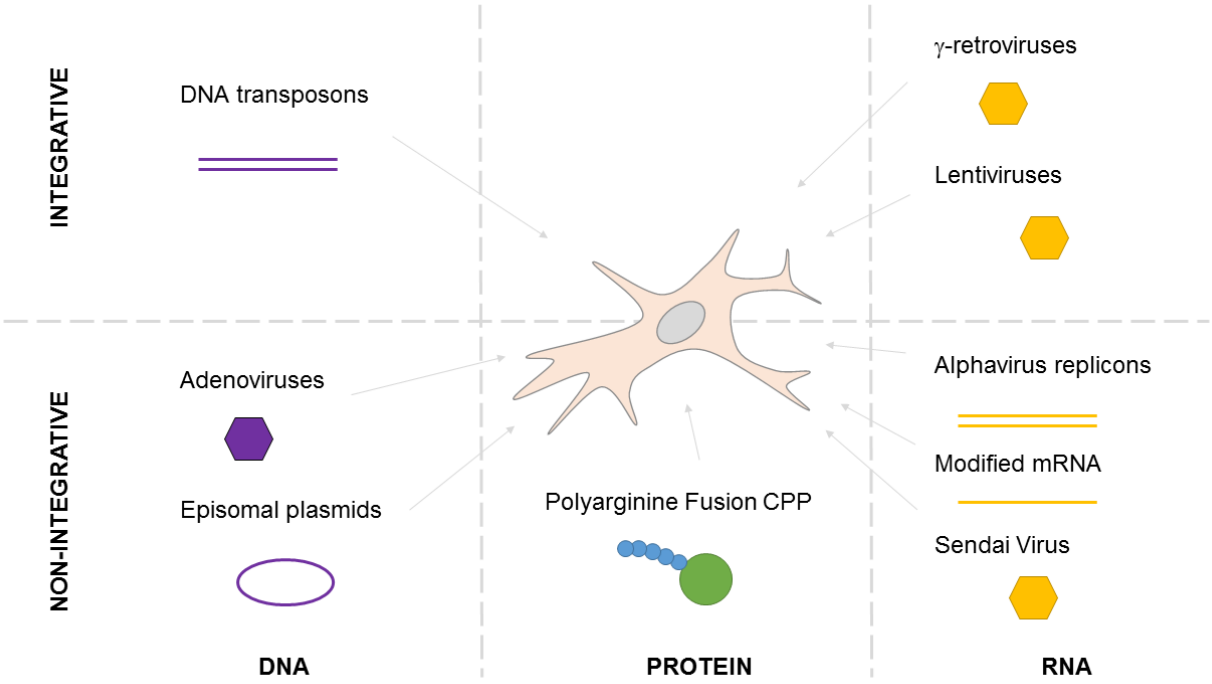

Figure 5. *Summary of cellular reprogramming vectors.*

6.2.5.1. Retroviruses

Retroviral vectors were used for the first iPSC induction experiments [1]. The genome of retroviruses consists of RNA which is reverse transcribed into DNA and integrated into the host genome upon infection. This allows for a persistent expression of the delivered transgenes. The retroviral gene transfer vectors have had most of the viral components removed and only contain the essential viral elements required for gene expression and delivery, including the ψ packaging signal, primer binding sequence, polypurine tract, and long terminal repeats (LTR) [344]. The retroviral gene transfer systems most commonly used for iPSC induction are primarily based on either mouse γ-retroviral vectors derived from the Moloney Murine Leukemia Virus (MMLV) or Lentiviral vectors derived from Human Immunodeficiency Virus [345,346]. These vectors have different properties when it comes to infectivity and expression. γ-retroviral vectors are ecotropic and normally only

infect mouse cells. These viruses can, however, be pseudotyped with the vesicular stomatitis virus G protein (VSV-G) which increases their infectivity to wide variety of cells. The γ-retroviral vectors require active cell proliferation for successful infection. Lentiviruses, on the other hand, can infect also non-dividing cells. Most γ-retroviral vectors used for iPSC induction rely on the viruses endogenous LTR promoter for gene expression, and get silenced towards the end of the reprogramming process by ZFP809, TRIM28 and CHAF1A mediated mechanisms [347]. Lentiviruses have commonly been engineered to contain exogenous promoters, which allow more control over their expression and for example the use of inducible promoters for secondary induction systems [348]. The problem with using retrovirus-based gene transfer methods for iPSC induction is the uncontrolled integration of the vector and the effects of the insertional mutations. It has been reported that high multiplicity of infection (MOI) with lentiviruses can mediate reprogramming of fibroblasts to pluripotency in the absence of transgenic transcription factors [349]. This has been linked to aberrant gene and miRNA expression caused by the vector integration.

6.2.5.2. Non-integrative Plasmids

Induction of pluripotency can be achieved with transient plasmid DNA transduction based reprogramming factor delivery methods. The first reports of this required either serial plasmid transfections or transfection of more plastic cell types, like adipose stem cells, due to the transient nature of plasmid expression [350,351]. This also generally resulted in inefficient reprogramming. The need for serial plasmid transfections was removed by introduction of replicative episomal plasmid vectors for reprogramming, which allowed for more persistent reprogramming factor expression until the endogenous pluripotency transcriptional program was properly established [352]. These vectors contain the Epstein-Barr virus derived oriP and EBNA1 (Epstein-Barr nuclear antigen-1) sequences, which mediate the replication of the plasmids once in a cell cycle. The EBNA1 protein recognises the oriP sequence and recruits the cellular DNA replication machinery to the origin. This system has later been further optimized for reprogramming factor composition to mediate more robust reprogramming of primary human cells [353]. As the reprogramming vectors are transiently lost during cell proliferation due to imperfect plasmid retention [344], the system allows for derivation of iPSCs without genomic vector integration.

6.2.5.3. DNA Transposons

In addition to transiently replicating episomal plasmids, integrative plasmid systems, based on DNA transposons, have been used for reprogramming. These systems consist of a vector donor plasmid, containing the transgenic cargo flanked by the terminal repeat sequences of the transposon, and a second helper plasmid expressing the transposase. Most commonly used DNA transposon reprogramming systems are based on the moth derived PiggyBac (PB) transposon, which efficiently transposes in PSCs [354,355], and can

mediated reprogramming of both mouse and human cells [356,357]. Additionally, DNA transposon vectors based on the Sleeping Beauty (SB) transposon, engineered from ancestral Tc1/mariner transposable elements from fish genomes, have been used for reprogramming [358,359]. As these vectors are integrated in the host genome, they can provide stable expression of the delivered transgenes. They can also be combined with inducible systems to control the temporal expression of the transgenes and to mediate secondary reprogramming of differentiated cells [356]. Compared to retroviral vectors, DNA transposons have higher cargo capacity, making them useful for delivery of larger constructs. SB and PB transposons are based on a cut-and-paste transposition model. This allows the removal of the reprogramming transgenes from the recipient genomes upon re-expression of the transposase protein. PB transposons integrate into the host genome at TTAA sites and can be precisely excised from the genome enabling the generation of transposition footprint-free iPSCs [360].

6.2.5.4. RNA Mediated Reprogramming

Modified mRNA mediated reprogramming relies on serial transfection of chemically modified mRNA molecules encoding the pluripotent reprogramming factors into the reprogramming cells [361]. Unlike transgenic DNA, transduced mRNA does not risk integration into the host genome. However, transfection of unmodified RNA generally causes an antiviral innate immune response in the cells resulting in increased interferon production and growth inhibition. The unmodified RNA is detected by RIG-I, PKR and TRL7/8, which can further lead to MAVS-mediated caspase activation and apoptosis [362–365]. Therefore efficient mRNA transfection requires inhibition of the innate immune response, which can be mediated by targeting the innate immune response factors for downregulation or inhibition [366,367], or by preventing the mRNA recognition by removal of terminal 5'-phosphates and introduction of modified ribonucleoside bases, such as 5-methylcytidine and pseudouridine [361]. These modifications, along with stabilizing UTR sequences, additionally increase the mRNA half-life, allowing for longer expression of the transduced mRNA and less frequent transfections. The mRNA transfection system is still relatively labor intensive method, requiring daily transfections of mRNA for approximately two weeks for proper reprogramming of fibroblasts. The system is also not quite as robust as the other commonly used non-integrative methods, episomal plasmid transfection and Sendai viral transduction [368]. The mRNA reprogramming efficiency can, however, be improved by simultaneous delivery of reprogramming promoting miRNAs [368]. These miRNAs are usually pluripotency associated micro RNA, like miR200c, miR302a-d, miR367, miR369 or miR372 [225,226,369–372]. In addition to promoting reprogramming, knock out of the miR302 cluster abolishes pluripotent reprogramming, demonstrating the importance of these miRNAs for the process [373]. The targets of these miRNAs in reprogramming include factors involved in cell cycle control, glycolytic metabolism, MET and gene repression [371].

6.2.5.5. RNA Episomes

Episomal RNA replicon vectors are based on alphavirus genomes. Alphaviruses are single stranded, positive sense, RNA viruses with 5'-capped and 3'-polyadenylated genome. Alphaviruses replicate via a double stranded RNA intermediate, which is polymerised by the viral non-structural proteins. The (-) sense RNA functions as a template for production of (+) sense full length genomic RNA and a shorter sub genomic (+) sense RNA, produced from an internal promoter. Alphavirus genomes contain two open reading frames translated from the genomic and sub-genomic RNAs. The first ORF encodes non-structural proteins 1-4, which are cleaved from a longer peptide, and the second ORF contains the viral structural proteins, which are important for producing virions [374]. As the structural proteins in the second ORF are not important for the virus expression or replication, they can be replaced with transgenic cargo. The alphavirus replicons are normally highly cytopathic, leading to apoptotic cell death in a few days, and the transgene vectors normally need to contain mutations, which can promote persistent maintenance of viral replicons. Multiple different alphavirus based transgenic replicons have been reported, including Sindbis Virus (SIN) [375,376], Semliki Forest Virus (SFV) [377], Chikungunya Virus (CHIKV) [378], and Venezuelan Equine Enchephalitis Virus (VEEV) [379]. Many of the alphavirus replicons are still highly cytopathic, limiting in their host cell options, and only persist in selected cell lines like BHK-21. The cytopathic effect is caused by the cells innate immunity factors, like RIG-I, MDA5 and PKR recognizing the viral double stranded RNA and initiating interferon response, MAVS-mediated apoptotic signalling and viral factor-mediated translational shut off [380–385] However, some new world alphavirus replicons, like VEEV, can be persistently maintained in various cell lines [379]. These vectors have turned out to be useful also for cellular reprogramming [386]. By replacing the viral structural proteins in the sub-genomic RNA with OCT4, SOX2, KLF4 and C-MYC encoding cassette, the vector can be used to derive iPSCs from human fibroblasts [386]. This way the (+) sense RNA replicon can be introduced by transfection or electroporation into cells, where it functions as an mRNA template for translation, and initiates replication of the delivered RNA construct and subsequent expression of the reprogramming transgenes. The double-stranded RNA is still toxic to the cells, which necessitates interferon inhibition for persistent maintenance of the replicon. On the other hand, removal of the interferon inhibition can be used for selection of vector-free clones by innate immunity response mediated removal of replicon containing iPSCs. As the whole system is based on RNA, it does not pose a risk for genomic integration.

6.2.5.6. Protein Transduction

Recombinant protein transduction relies on the fusion of the cargo proteins with a cell penetrating peptide (CPP), which allows for the translocation of the protein across the cell membrane from extracellular space into the cell. The CPPs are short peptide sequences, which can be divided into three categories, cationic, amphipathic and hydrophobic, based on the physical properties of their amino acid sequences [387]. Cationic peptides, like HIV-TAT peptide or polyarginines, contain multiple positively charged residues, like lysines and arginines, which promote cell entry by interacting with negatively charged groups of cell membrane constituents. Many amphipathic CPPs are chimeric peptides between hydrophobic domain and a cationic nuclear localization signal, and hydrophobic CPPs consist of nonpolar residues, which have high affinity for hydrophobic domains of cell membranes [387]. CPPs can exploit different mechanisms for cell entry, but most CPPS are internalized by endocytosis followed by escape from endosomes. This process can be promoted for example by addition of extra histidine residues to the peptide tag, which adsorb protons at endosomal pH leading to increase in osmotic pressure and endosomal vesicle membrane rupture [388]. Pluripotent reprogramming attempts have been reported with both TAT and polyarginine fused reprogramming factors. Although the reported TAT fusion factors did not translocate properly in the cells, transduction with polyarginine fused factors can mediate reprogramming of both mouse and human fibroblasts [389–391]. Direct protein transduction of reprogramming factors is still an inefficient method. The first reported human iPSC derived by protein transduction were derived from fibroblasts with abnormal karyotypes [390]. Later advances have improved the protein transduction reprogramming approach by including an innate immunity signalling activating dsRNA analog (poly- inosinic-polycytidylic acid), which promotes polyarginine fused protein-mediated reprogramming [392]. Additionally, less specified somatic cell pluripotent reprogramming approaches have been reported by using streptolysin O–mediated reversible permeabilization with ES cell protein extract transduction [393].

6.2.5.7. Episomal Viruses

In addition to the integrative retroviral approaches, various non-integrative virus vectors have been used for induction of pluripotency. These include both episomal DNA and RNA viruses, like adenovirus or Sendai virus (SeV). Adenoviral vectors can have large cargo capacity if all viral proteins are removed and only necessary packaging signals are left. The vector carries a linear double stranded DNA molecule in the cells and provides a transient expression in proliferating cells [344]. Adenoviral reprogramming factor delivery has been used to derive integration free mouse iPSCs, but the efficiency of the system is low [394]. Sendai virus is a single stranded negative sense RNA virus capable of efficient transduction of various cell types. Sendai viral proteins NP, P and L form a ribonucleoprotein complex with the viral genome and mediate its replication in the cytoplasm. The other proteins, M, HN and F, mediate the formation of virus particles and viral infection. F-defective

Sendai viruses are incapable of producing infective viruses and have thus been developed as gene transfer vectors [344,395]. Due to the episomal replication of SeV genomes, SeV vectors are capable of persistent expression of transgenes and do not require repeated transduction. This makes SeV vectors attractive tools for pluripotent reprogramming. F-defective SeV vectors have been used to reprogram somatic cells [396]. The vector provides high transgene expression levels and can mediate efficient reprogramming. The replicons tend to be lost over time, but they can persist quite long. The clearance of the transgenic replicons can be promoted by using temperature sensitive mutants of the virus and transfer of the infected cells to higher temperature [397]. Sendai viral reprogramming has also been reported using an alternative virus strain capable of persistent long term replication [398]. This vector has additional M and HN protein deletions, and can incorporate all reprogramming transgenes in a single cassette, providing more control over factor expression stoichiometry. The clearance of the vector can be mediated either by transfection with siRNA targeting the SeV genome or by incorporating miR302 recognition sites in the vector sequence [399]. This provides more efficient clearance of the vectors than passive loss of viral replicons [400].

6.3. Adeno Associated Virus (AAV)

Adeno-associated virus (AAV) is a small non-pathogenic parvovirus that is dependent on a helper virus, such as adenovirus or herpesvirus, for productive infection. It was initially found contaminating Adenoviral preparations and has later on been developed as a gene transfer vector [401,402]. AAV genome consists of a 4.7 kb single stranded DNA, containing three promoters and two open reading frames flanked by hairpin forming inverted terminal repeat (ITR) sequences. The left half contains the Rep gene encoding the non-structural proteins Rep78, Rep68, Rep52, and Rep40, which bind to the Rep binding elements in the ITR and exhibit DNA helicase and endonuclease activity at the ITR terminal resolution site [403,404]. The Rep proteins control the AAV gene expression in response to the helper virus and are required for AAV DNA replication and accumulation of single stranded viral DNA for packaging [404,405]. The right half contains the Cap gene producing the structural proteins, VP1, VP2 and VP3, which form the viral capsid. Additionally, the AAV replication is dependent on helper virus factors, like the adenoviral E1, E2A, E4orf6 proteins and VA1 RNA [402]. Upon infection, AAV can establish a latent state in the absence of helper virus when conditions are not permissive for productive replication. Infection of the latent AAV containing cells with a helper virus can re-establish infective AAV production [402]. The latency of the AAV provirus can be established by circularization of either singular or concatenate AAV genomes into extrachromosomal episomes in non-dividing cells or by integration into the host genome. The integration of the wild type AAV happens preferably into the AAVS1 locus in human chromosome 19q13.3-qter and is mediated by the viral Rep proteins [406,407]. This is due to the presence of a Rep binding element sequence in close proximity to a terminal resolution site in the AAVS1 locus, rep binding elements in the AAV ITR sequences and p5 promoter, and tethering of AAVS1 and AAV genomes by Rep68 [404]. The exact integration site itself is imprecise. The ITRs are approximately 145 bp long sequences, containing the AAV replication origin, and are required for the viral genome packaging into the virions. After infection and viral DNA entry into the host cell the ITRs mediate the second strand DNA synthesis by self-priming [404], and the circularization of the AAV episomes by DNA damage repair [408]. AAV enters into the cells via cell surface receptors, including the recently described AAVR [409], and gets internalised into endosomal vesicles, which it exits upon acidification. AAV is actively transported into the nucleus, which allows AAV-mediated transduction of non-dividing cells. Cellular processes that mediate AAV entry, trafficking and viral uncoating can limit AAV transduction efficiency [410]. Different serotypes of AAV can mediate these processes to varying extent. At least 13 different AAV serotypes have been described so far, which demonstrate varying degree of tissue tropism *in vitro* and *in vivo* [411,412].

6.3.1. Recombinant AAV (rAAV) as a Gene Delivery Vector

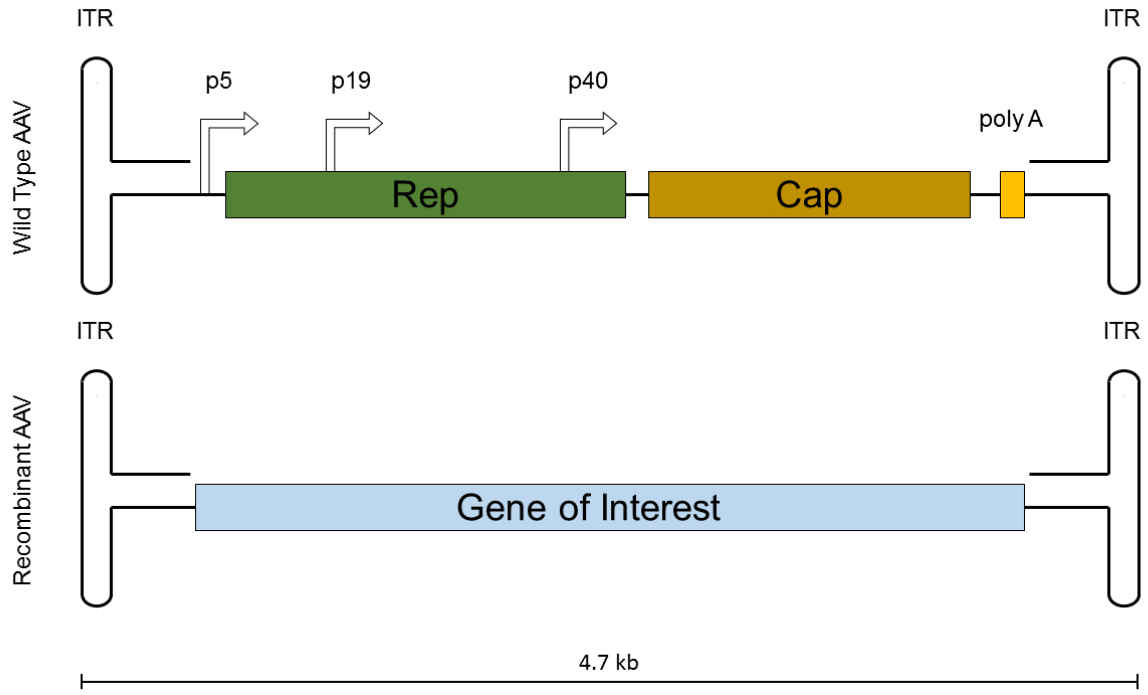

Figure 6. *Schematic representation of AAV genome structure.*

The fact that AAV is a non-pathogenic and non-toxic virus, capable of long term persistent tissue specific transduction of non-dividing cells, makes it an attractive tool for a gene transfer vector, especially from a gene therapy point of view. To this extent, gutless recombinant AAV vectors have been developed, which lack all of the viral genes. As the ITR sequences mediate both the genomic DNA packaging to the virions and the second strand DNA synthesis priming in the transduced cells, they are the only viral sequences required for rAAV gene transfer vectors. Therefore, recombinant AAV vectors can be generated by flanking a transgene with ITR sequences and providing the AAV proteins and helper virus proteins in trans for virus production and packaging. As the wild type AAV genome is relatively small, the carrying capacity of rAAV vectors is also quite limited. Maximal capacity of 5.2 kb has been reported to be the upper limit for rAAV packaging [413], but longer constructs may work by annealing or recombination of truncated vector sequences or by trans-splicing between two vector constructs [414]. Small genes can also be packaged into self-complementary rAAV vectors. Self-complementary rAAV vectors can fold back onto themselves removing the need for the rate limiting step of second strand DNA synthesis [415–417]. The easy pseudotyping of rAAV vectors with different capsid proteins and the variety of available AAV serotypes makes possible to engineer rAAV with desired target tissue specificity. Moreover, targeted capsid tyrosine mutations, DNA family

shuffling and *in silico* reconstruction of ancestral AAV sequences can yield highly improved rAAV gene transfer vectors [418–420]. Schematic representation of AAV genome structure is presented in **Figure 6**.

As the rAAV is lacking the viral Rep proteins it cannot meditate site specific targeting to the chromosome 19 AAVS1 locus. Therefore the rAAV vectors preferentially tend to form monomeric and concatemeric circular episomes, as has been reported for mouse muscle tissue transduced rAAV vectors [421]. rAAV vectors can however integrate into the host genome, but appear to do so in a more opportunistic way by utilising genomic DNA double strand breaks (DSB) [422]. rAAV integration *in vivo* has also been reported in the male germ cell lineage [423]. The integration of rAAV genomes appears to have a preference for expressed genes [424,425], and therefore it is possible that the rAAV integration is liked to transcription-associated DNA DSBs [249]. As AAV genome processing co-localizes with DNA damage response factors of the MRN complex (MDC11-RAD50-NBS1) [426], it is possible that the AAV genome processing is also linked to the integration mechanisms and that the integration to DNA DSB sites occurs by microhomology mediated alt-NHEJ pathway, as suggested by the vector integration junction sequences [422,427]. The integration of rAAV vectors is still a relatively rare occurrence and the integration frequency has been estimated to be in the range of 0.2-1.0 × 10^{-3} per infectious vector genome [428]. This makes rAAV vector a potentially useful tool for applications that require transient expression in replicating cells and a subsequent absence of integrated vector sequences, such as iPSC induction.

6.4. CRISPR/Cas Systems

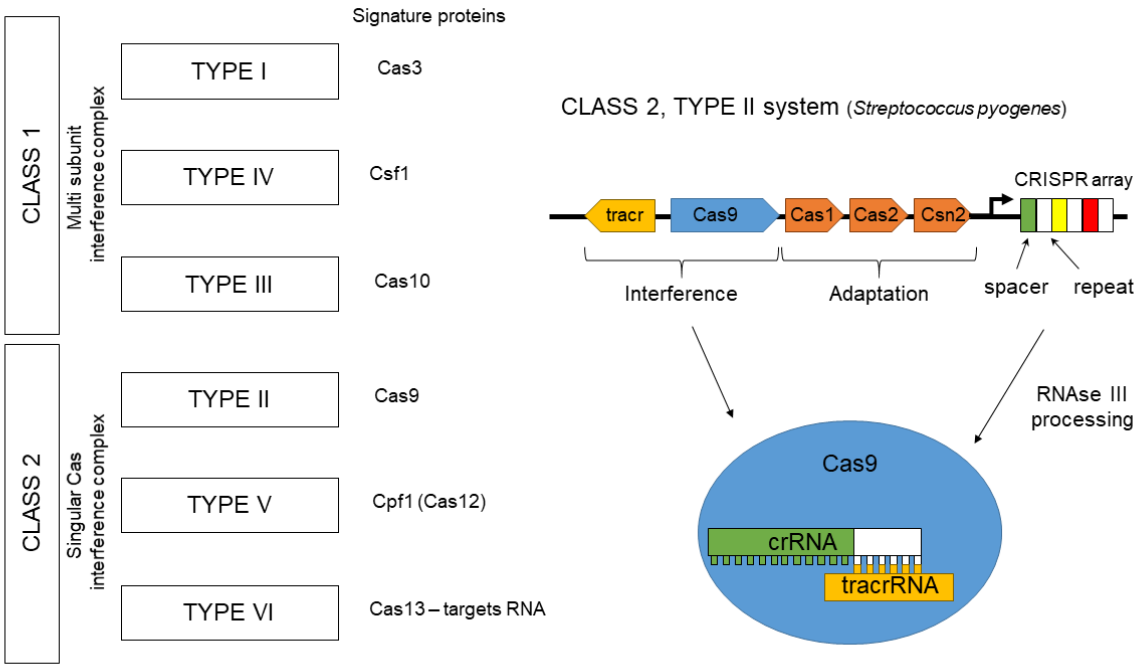

Figure 7. *Classification of CRISPR-Cas systems and schematic representation of Class 2 Type II CRISPR system of Streptococcus pyogenes. Genome editing tools are primarily based on single Cas interference complex containing Class 2 Type II and V systems. Streptococcus pyogenes CRISPR locus contains Cas1, Cas2 and Csn2 proteins that mediate new spacer acquisition in the CRISPR array, and Cas9, tracrRNA and CRISPR array that mediate interference of invading genetic material.*

The research of clustered regularly interspaced short palindromic repeats (CRISPR) sequences traces its origin to the study of repeat motifs in prokaryotes and archaea in the late eighties and early nineties [429]. Short repetitive sequences containing variable spacer regions were initially described in various bacteria, including *E. coli,* and halophilic archaea [430–433]. These repeats were later named CRISPR and found to contain nearby protein coding genes named CRISPR-associated (Cas) [434]. The function of these repeat regions was not initially understood, but discovery of matching spacer sequences in bacteriophages and plasmids implicated a potential role of these loci in bacterial immunity against foreign DNA [435]. The empirical evidence for CRISPR function in bacterial adaptive immunity was demonstrated in 2007 when a phage sensitive strain of *Streptococcus thermophilus* was shown to incorporate phage protospacer sequences into its CRISPR1 locus upon acquisition of phage-resistance after phage challenge [436]. Moreover, removal of the spacers from the resistant strain re-sensitized the bacteria to the phage infection, and replacement of the wild type CRISPR1 locus with phage specific spacer containing sequence-conferred resistance to the phage. The role of CRISPR RNAs in guiding the immune response and the mechanism of target DNA cleavage were subsequently

demonstrated [437,438], followed by the requirement for a trans-encoded CRISPR RNA (tracrRNA) for *S. pyogenes* phage resistance [439]. The characterization of the necessary components eventually allowed reconstitution of the RNA reprogrammable Cas9 endonuclease function from its minimal constituents [3,4]. This enabled the further development of the system as a genome-engineering tool and lead to the rapid expansion of CRISPR research and applications.

CRISPR-Cas is a widely spread adaptive immunity system present in approximately 47 % of bacterial and archaeal genomes. CRISPR systems can be classified into two classes, based on the presence of either multi subunit (class 1) or single protein (class 2) interference complexes, and subsequently into six types based on the presence of unique signature proteins, Cas3 (type I), Cas9 (type II), Cas10 (type III), Csf1 (type IV), Cpf1 (Cas12) (type V) and the RNA targeting Cas13 (type VI) [440,441]. In CRISPR-Cas immunity, the first step is adaptation to infection by insertion of invading genetic material into the CRISPR locus by Cas1 and 2 [442], which is promoted by defective phage infection [443]. Thereafter, the CRISPR locus is expressed and the precursor CRISPR RNA (crRNA) are bound to the multi subunit or singular Cas interference complexes and processed. In the final step the interference complexes mediate target sequence recognition and cleavage [440]. The primarily bacterial class 2 single protein interference complex systems have been a major focus for the biotechnological development of CRISPR-based applications due to their relative simplicity. Particularly class 2 type II Cas9 proteins and class 2 type V Cpf1 proteins have been developed as genome editing tools. These systems in bacteria usually consist of a CRISPR locus containing the crRNA sequences, Cas1 and Cas2 containing adaptive complex, Cas9 or Cpf1 containing interference complex, and in case of type II systems tracrRNA, which is needed for proper crRNA targeting [3,4,440]. The crRNA processing additionally requires RNAse III. The crRNA, or guide RNA, consists of a repeat sequence, which mediates Cas interaction, and a spacer sequence, which defines the genomic target specificity. The spacer sequences are derived from protospacer sequences that usually originate form invading genetic material. The CRISPR/Cas9 system requires a tracrRNA [3,4], whereas CRISPR/Cpf1 system is dependent on only one guide RNA molecule, which is processed by the intrinsic RNase activity of Cpf1 [444,445]. For DNA targeting, the Cas protein needs a protospacer adjacent motif (PAM) next to the guide target site. PAM sequences are specific to different Cas proteins and are recognised by the Cas RNP complex in a spacer sequence independent manner in the first step of target site binding [446]. This is followed by sequence specific spacer RNA annealing to the target DNA leading to a conformational change in the Cas9 protein, which promotes catalytic activation of the protein [447]. In case of Cas9 the cleavage is mediated by nicking of each DNA strand by the HNH and RuvC domains of the protein, producing a blunt end double strand cut [4], whereas Cpf1 cuts the DNA strands sequentially and produces a staggered cut with 5' overhangs [444]. A schematic representation of CRISPR systems is presented in **Figure 7.**

6.4.1. CRISPR/Cas9 as a Genome Editing Tool

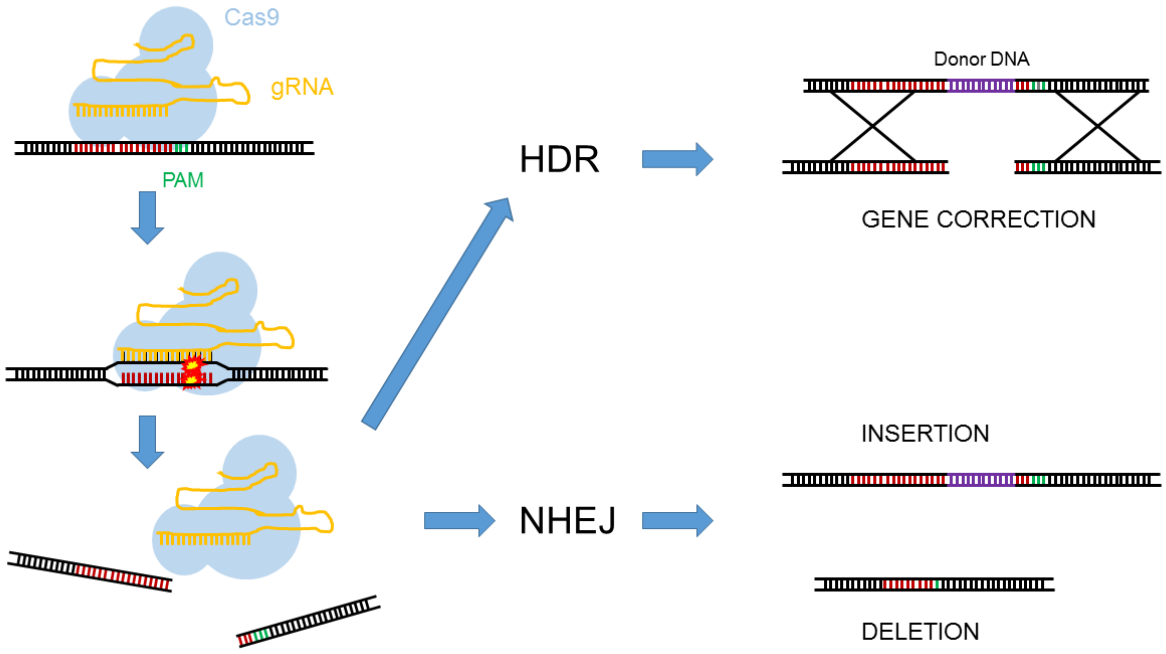

Figure 8. *Basic principle of CRISPR/Cas9-mediated gene editing. A programmable nuclease is introduced to a cell to generate a site-specific DNA double strand break. Thereafter the cell intrinsic DNA repair mechanisms correct the DNA double strand break using either homology directed repair (HDR) or non-homologous end joining (NHEJ). NHEJ mediated error prone repair can produce insertions and deletions that interfere with the original sequence. HDR can be used to introduce donor DNA cassettes into the targeted site.*

CRISPR/Cas9 has attracted lot of attention as a potent genome editing tool. Genome editing with programmable nucleases relies on targeted introduction of a DNA double strand break, which will be corrected by the cells intrinsic DNA repair machinery primarily via either recombination by homology directed repair (HDR) or non-homologous end joining (NHEJ). These targeted DNA DSBs can be used to produce mutations, such as small insertions and deletions, which can interfere with open reading frames, or to increase the rate of HDR occurring at the breakage site [448,449]. First genome editing approaches relied on either meganucleases, zinc-finger nucleases (ZFN) or transcription activator-like effector nucleases (TALEN). Meganucleases are restriction enzymes, which have large recognition sites that are generally uncommon in the genome, making it possible to target the genomic sites with some specificity [449]. However, the recognition sites are protein specific and the proteins are difficult to engineer, making the use of meganucleases impractical for genome editing. ZNF are fusions of zinc-finger domains of DNA binding proteins to a catalytic domain of FokI endonuclease. Zinc-finger domains recognise DNA sequences in triplets and fusion of multiple zinc-finger domains can be used to extend the target site sequence of the constructs. ZFNs are used as pairs

and recruited to two recognition motifs spaced 5-7bp apart. This promotes FokI dimerization in the middle and DNA cleavage at the site [448,450]. ZFNs can be reprogrammed by modular assembly of the domains, but construction of functional high activity ZFNs is generally still a quite challenging and lengthy process. More efficient generation of modular reprogrammable nuclease proteins can be achieved using TALENs. TAL effectors originate from the plant pathogenic *Xanthomonas* bacteria and are naturally modular DNA binding proteins [451]. TAL effector DNA binding domain monomers recognise single bases in DNA and can be combined together for longer target sequences. This can be used to mediate DNA DSB generation by paired TALEs fused with FokI domains, similar to ZFNs, or to mediate endogenous gene activation with activator domain fusions [452,453]. Like ZFNs, TALENs also require extensive protein engineering to reprogram their target site specificity, which may complicate their use.

As the class 2 CRISPR-Cas interference systems only require one protein component for DNA cutting, these systems are good candidates for genome engineering tools. Particularly the type II system Cas9 proteins have been utilized for this approach. As the Cas9 targeting is achieved with two short RNA molecules, or one single guide RNA with fused crRNA and tracrRNA molecules, the system does not need protein engineering for re-targeting [3,4]. This makes it easier and faster to target new sites for genome engineering with CRISPR/Cas9 than with ZFNs or TALENs. By incorporating a nuclear localization signal and relevant expression control elements into the *Streptococcus pyogenes* Cas9 protein, this system can be used to edit human genes [454,455]. Moreover, as the Cas9 protein is targeted to DNA by short guide RNA molecules, it is easy to simultaneously target the system to multiple sites using one constant protein component and multiple guide RNA molecules [455]. Basic principle of CRISPR/Cas9-mediated gene editing is presented in **Figure 8**.

6.4.1.1. SpCas9

The most commonly used CRISPR system for human cell engineering is based on the *Streptococcus pyogenes* Cas9 (SpCas9) and guides. SpCas9 normally binds two RNA molecules, but in biotechnological applications these are normally fused together with a hairpin loop to mediate easier expression [4]. The Cas9 protein contains two nuclease domains, HNH and RuvC, which each nick one strand of DNA, protospacer complementary strand and protospacer non-complementary strand respectively, producing DNA DSB. These sites can be inactivated by incorporating point mutations into the catalytic sites of the domains, the most commonly used mutations being D10A in the RuvC domain and H840A in the HNH domain [4]. Incorporating only one of these mutations causes the Cas9 to have nicking activity, cutting only one strand of DNA, which can be a useful way of reducing off target mutagenesis as no DNA DSBs are created by default. By combining two guides with a nicking Cas9 protein, the off target mutagenicity of the system can be reduced from 50- to 1500-fold, as creation of DSBs will require two closely spaced target sites [456]. As the guide RNA starts to bind

DNA from the PAM proximal sequences, mismatches close to the PAM are generally less tolerated than mismatches further away from the PAM, especially close to the 5' end of the RNA [457,458]. Additionally, the guide nucleotide composition may affect the effectivity of the guides, with certain bases preferred over others in different parts of the guide sequence [459]. Enhanced specificity versions of the protein have been rationally designed based on the crystal structure of the SpCas9 protein [460,461]. These versions reduce the Cas9 affinity with the non-target DNA strand (eSpCas9) or the target DNA strand (SpCas9-HF1), thus decreasing the affinity of the protein to mismatched sites [462,463]. More recently, additional high specificity versions of Cas9 have been reported, with mutations primarily in the REC3 lobe. One of these variants has mutations affecting the threshold for conformational shift required for the Cas9 catalytic activation [464], and another one was identified in an *in vivo* yeast screen for Cas9 specificity [465].

The SpCas9 is generally considered to recognise the consensus DNA sequence NGG as a PAM motif. The recognition of the motif is however dependent on the surrounding nucleotide sequences and variants of this motif have been characterised, often with an A in the second position [466]. As the PAM specificity is defined by the Cas9 protein, targeting of PAM variants that differ from the SpCas9 PAM consensus sequence require modifications in the Cas9 protein or the use of orthogonal Cas9 or Cpf1 systems [467,468]. Reported modified SpCas9 variants can recognise PAM variants NGA, NGAG and NGCG instead of the wild type NGG consensus [469–471], or more broad PAM sequences including NG, GAA and GAT [472].

Incorporating inactivating mutations in both of the catalytic domains of the Cas9 protein, it can be converted in to a dead Cas9 (dCas9) version which does not cut DNA upon binding. This can be utilised as a reprogrammable DNA binding protein and a scaffold with which other functional components can be recruited to specific genomic loci [473]. The use of trans-activator fusions and epigenetic dCas9 effectors will be discussed later. Other mentionable uses include dCas9-mediated induced chromosome looping [474,475], visualization of DNA loci in live cells [476–478], enrichment of chromosome locus interacting factors [479–481], and DNA base editing [482]. DNA base editing is of particular interest for genome editing as it directly converts DNA bases in the presence of the Cas9 binding site without the need for a DNA DSB. Nicking activity of the Cas9 protein can, however, help bias editing to the right DNA strand. Base editing relies on the fusion of nicking Cas9 with APOBEC1 cytidine deaminase, for C to T conversions, or evolved tRNA adenosine deaminase, for A to G or conversions [482,483]. Fusion of these components with orthogonal Cas9 proteins and SpCas9 PAM variants can further expand the targeting range of these tools [484].

6.4.1.2. Orthogonal Class 2 CRISPR Systems

Although *Streptococcus pyogenes* CRISPR/Cas9 system is most commonly used for mammalian genome engineering, many other orthogonal CRISPR systems have been characterised. The main benefit of using

orthogonal Cas9 systems comes from the varying PAM specificities of the proteins, which help expand the genomic targeting range of the method. Frequently used alternative class 2 CRISPR systems include *Neisseria meningitidis* and *Staphylococcus aureus* Cas9 [485,486], and *Acidominococcus* and *Lachnospiraceae* Cpf1 [444]. Orthogonal CRISPR systems may be useful in applications where multiple genomic loci need to be affected simultaneously with different proteins, since the different Cas9 and Cpf1 proteins have their own specific guide RNA structures, allowing designs with minimal guide RNA cross recognition [487]. Additional benefits may include smaller size of the Cas9 proteins, as in the case of *Staphylococcus aureus* Cas9, allowing the use of the system in limited capacity vectors like rAAV [486], or increased thermal stability of the *Geobacillus stearothermophilus* Cas9, which promotes higher stability *in vivo* and use in higher temperature biotechnological applications [488].

6.4.1.3. Guide RNAs

In bacteria, the CRISPR locus is expressed in a single polycistronic transcript, which is processed into individual crRNAs and bind to Cas9 with the tracrRNA [439]. Separate crRNA and tracrRNA molecules work in mammalian cells but, to mediate more efficient utilisation of this system, the two small RNAs are commonly fused into one single guide RNA molecule (sgRNA) with a short hairpin loop [4,455]. The guide RNAs are usually expressed under ubiquitously expressed Pol III promoters, such as U6 or H1, with specified start nucleotide and terminator sequences. Multiplex guide assembly is commonly achieved by Golden Gate cloning, which allows rapid concatenation of modular repetitive sequence motifs, such as the guide expression cassettes, into single vectors [489–492]. Other means for guide concatenation include expression from a single transcript and subsequent processing by incorporation of external RNA endonuclease sites [493], tRNA promoters and intervening tRNA sequences [494–496], self-cleaving ribozymes [497,498], and intervening miRNA or shRNA processing by DROSHA [499,500]. On the other hand, the use of Cpf1 allows direct concatenation of the guide sequences, due to the intrinsic RNA processing capacity of the enzyme [445], and more straight-forward guide multiplexing without intervening sequences or tracrRNA [501–504]. Additional guide mediated Cas9 functionalisation can be achieved by shortening the guide RNA molecules so that the guide 5' end binding dependent Cas9 conformational change does not take place. This prevents the wild type Cas9 from cutting DNA properly at the target site and allows the use of Cas9 instead of dCas9 for site specific effector delivery [505,506].

6.4.2. CRISPR-Mediated Transcriptional Control

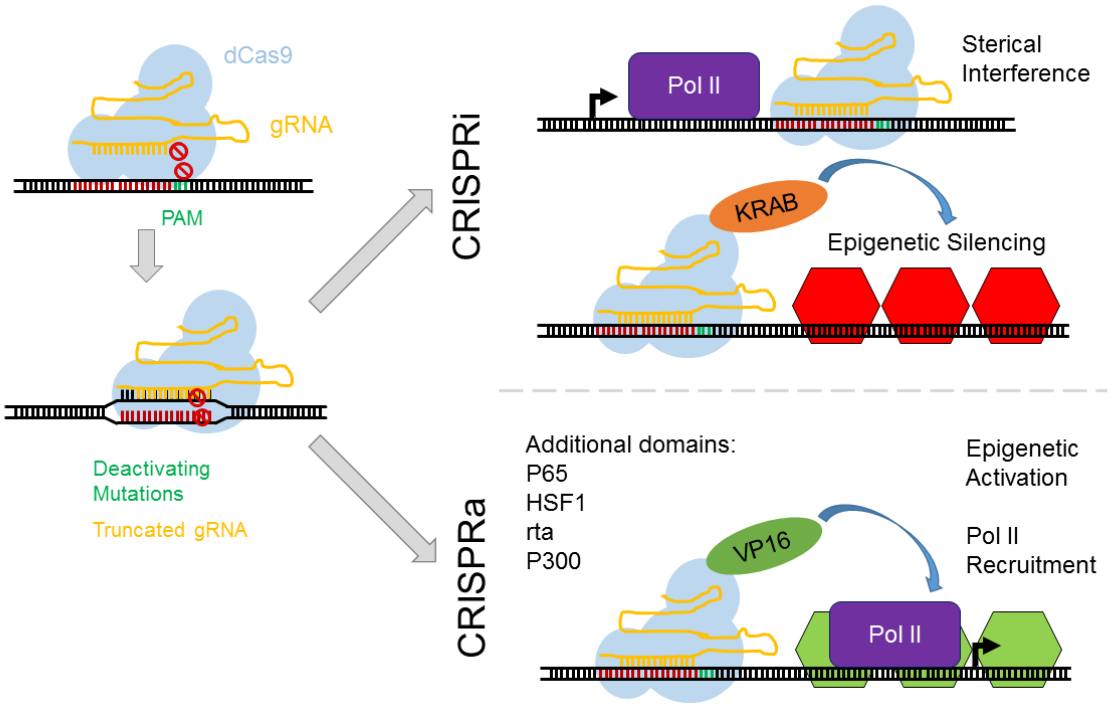

Figure 9. *Basic principle of CRISPR/Cas9-mediated transcriptional control. Cas9 enzymatic activity is deactivated by mutations in the catalytic domains or by truncation of the guide RNA to prevent conformational changes required for DNA cleavage. Targeting dCas9 downstream of transcription start site can be used to interfere with transcription machinery. Fusion of dCas9 with repressive domains can be used to epigenetically silence target genes. Fusion of dCas9 with transactivation domains can be used to epigenetically open target regions or to recruit RNA polymerase II to activate gene transcription.*

Cas9 or Cpf1 proteins, which lack DNA cutting activity, can be used for transcriptional control of genes. In its simplest form dCas9 recruited downstream of the transcription start site or in core promoter elements can be used to interfere with transcriptional elongation or RNA polymerase binding to reduce gene expression [507]. In the case of RNA elongation interference downstream of the TSS, the dCas9 protein needs to be targeted to the non-template DNA strand for proper inhibition, possibly due to more efficient physical interference with the polymerase. Various activator and inhibitor domain fused dCas9 constructs can also be used to control transcription in bacteria and in eukaryotes [508–514]. The most commonly used effector domains for gene regulation are Krüppel-associated box (KRAB) domains for gene repression and NF-kappa B p65 subunit trans-activation domain and multimeric repeats of the acidic activator peptide of the HSV1 VP16 trans-activator protein for gene activation [509,515]. These domains help recruit effector complexes to the dCas9 bound region. These are either repressive complexes, in case of KRAB, or various activator complexes, including basal transcription machinery and histone acetyl transferases, in case of p65 and VP16. dCas9 activators help

recruit the RNA Pol II to the gene promoter and to initiate transcription if the guide RNAs are targeted upstream of the TSS. Generally, recruiting multiple guides to a promoter or increasing the number of repeats of the trans-activator domains appears to increase gene activation potential [510–513]. The most drastic examples of multiple trans-activation domain recruitment to the dCas9 protein have been demonstrated by using multimeric peptide tags, such as SunTag, utilizing GCN4 antibody-peptide pair, or GFP11, utilizing the pairing of the 11th strand of the GFP β-barrel and the non-fluorescent GFP1-10 fragment [516,517]. In addition to fusing the activator domain directly with the Cas9 protein, the domain can be linked to a protein motif recognizing an RNA aptamer incorporated into the guide RNA molecule [513,518,519]. The benefit of this system is that by including multiple aptamers in the RNA, multiple activator domains can be recruited to one dCas9 RNP complex [518]. Additionally, various aptamer sequences, recognized by different protein motifs, can be used to define guide specific trans-activation effects without changing the dCas9 protein core [520]. Recruitment of multiple different trans-activation domains to the Cas9 protein can also help in improving the activation potential of dCas9 effectors. This has been reported either by recruiting P65 and the heat shock factor 1 (HSF1) transactivation domain to guide RNA aptamers by the MS2 motif [518], or by direct fusion of the p65 and the viral Rta transactivation domain to a dCas9VP64 core [521]. These transactivation domains are generally highly potent in endogenous gene activation and can even be used with single guide RNAs targeted per promoter. However, they appear to hit an efficiency threshold, as recruitment of multimeric repeats of these domains does not appear to further enhance their gene activation potential [522]. Basic principle of CRISPR/Cas9-mediated transcriptional control is presented in **Figure 9**.

6.4.2.1. Inducible CRISPR Systems

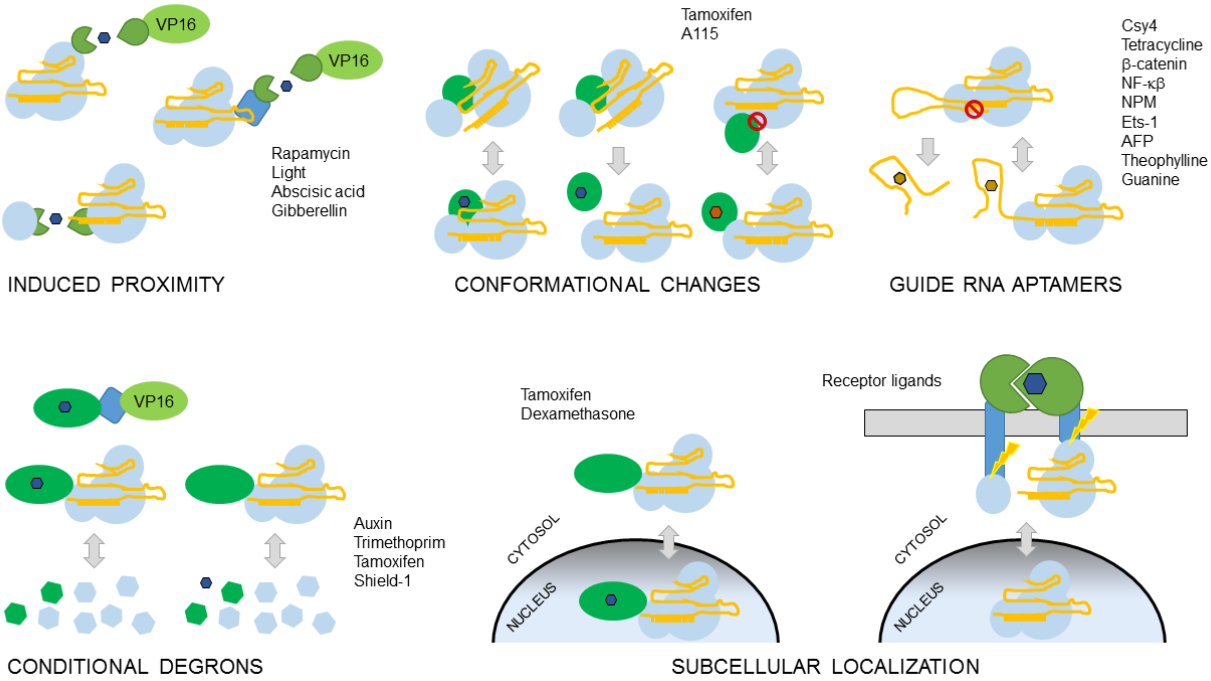

Figure 10. *Summary of inducible CRISPR systems.*

To gain more control over the activity of the CRISPR genome editing and transcriptional regulation, various inducible systems have been developed [523]. These systems can work either by controlling the Cas9/Cpf1 protein component or the guide RNAs. The simplest systems use inducible promoters to control the expression of the CRISPR transcripts. A commonly used method is to put the Cas9 protein under a doxycycline inducible promoter and insertion into the AAVS1 locus, alongside constantly expressed reverse tetracycline trans-activator [524,525]. Inducible guide RNAs can be engineered by incorporating ligand-responsive riboswitches into the guide RNA sequence. These systems rely on an antisense oligo sequence in the guide RNA that inhibits the guide binding to its DNA target site in the off-state. After ligand binding to the RNA aptamer, the molecule goes through either a cleavage or a conformational change that removes the antisense oligo binding to the guide sequence and releases the guide to bind to its target DNA. Riboswitch guide constructs have been reported for both Cas9 [526–528], and Cpf1 systems [529], with various input ligands, such as small molecular compounds and signalling factors. Protein mediated control of CRISPR/Cas9 activity can be achieved using multiple different means, including conditional degron fusions, induced proximity, inducible conformational changes, and subcellular localization control. These systems also use various different activating components, like external and cell intrinsic proteins, small molecular compounds and

light, enabling complex multimodal control over gene editing and transcriptional regulation [530]. Reported induced proximity methods have utilised, for example, rapamycin or blue light inducible split Cas9 architecture for genome editing [531,532]. For transcriptional and epigenetic control, light, abscisic acid, gibberellin or rapamycin have been used to induce binding of effector domains to dCas9 [533–537]. Inducible degrons rely either on default degradation and inducible stabilisation or inducible degradation of Cas9 or the activator domain. CRISPR/Cas9 systems utilising the inducible degradation approach have been reported with auxin, trimethoprim, tamoxifen and Shield-1 inducible systems [520,538,539]. Inducible conformational changes have been utilised with tamoxifen controllable allosteric changes and intein splicing [540,541], light inducible conformational changes [542], and chemically inducible interference of Cas9 intramolecular autoinhibition [543]. Subcellular localisation of Cas9 or split Cas9 has been controlled by ERT2 fusions and tamoxifen inducible nuclear import [544,545], and by glucocorticoid receptor fusion [546]. In addition to the external light and small molecular inducers, Cas9 and Cpf1 systems that are inducible by synthetic receptor interactions and signalling factor binding guide aptamers can be used to control the system activity in response to the cellular microenvironment [527,529,547–549]. Inducible CRISPR systems are summarised in **Figure 10**.

6.4.3. CRISPR-mediated Epigenome Editing

In addition to directly affecting gene transcription from promoters, the reprogrammable genome targeting tools have also been applied for modifying the epigenetic state of targeted loci [514]. This is commonly done either by fusing an interaction domain with dCas9 that can recruit chromatin modifying enzymes or by directly fusing the catalytically active domain of these enzymes to the dCas9. For example, the fusion of P300 catalytically active domain with dCas9 has been shown to affect target site H3K27 acetylation and activate genes from enhancers [550], and the fusion of dCas9 with human HDAC3 has been demonstrated to reduce H3K27 acetylation at its target sites [551]. Additionally, dCas9 fusion with the catalytic domain of the LSD1 histone demethylase has been demonstrated to cause enhancer position specific histone modifications and transcriptional changes that differ from those caused by the commonly used KRAB repressive domain [552]. Fusion of dCas9 with catalytic domains of DNA modifying enzymes have also been described for editing DNA methylation status. Particularly, DNMT3a catalytic domain fusion with dCas9 can mediate *de novo* methylation of CpG sites [553,554], and TET1 catalytic domain fusion with dCas9, or in combination with the SunTag system, can mediate targeted DNA demethylation [554,555].

In addition to dCas9 fused with effector domains, the Cas9 binding itself can cause changes in the local chromatin structure at the target site. The binding of the dCas9 protein in nucleosomal DNA can open the chromatin next to the dCas9 target site and promote binding of transcription factors to the target site proximal regions up to 100 bp form the guide binding site [556]. The effect can also be utilized to improve

specificity of DNA editing by promoting SpdCas9 mediated local chromatin opening in combination with Cas9 and Cpf1 orthologues, which by default target nucleosomal DNA with low efficiency [557].

6.4.4. CRISPR-mediated Control of Cell Fate

Due to the fact that CRISPR-mediated transcriptional control targets endogenous genes for activation or inhibition, the system appears useful for cell type conversions, such as reprogramming, where stable changes in endogenous gene expression programs are required. CRISPR-mediated endogenous gene targeting may help overcome reprogramming barriers, for example by directly targeting genomic loci which are normally not efficiently targeted by limited sets of reprogramming factors, or by targeting gene control regions that help maintain gene expression, such as enhancers. Additionally, as the CRISPR system relies on short gRNA molecules for defining the Cas9 binding sites, the system has a high gene targeting capacity with relatively small constructs. This means that instead of optimising gene regulatory network targeting for a minimal set of reprogramming factors, the targeting can be done with a more comprehensive approach aimed at maximising the number of targeted genes. The easy programmability of the Cas9 targeting allows the use of the CRISPR system for many types of cell fate control approaches, including reprogramming, transdifferentiation and differentiation. So far CRISPR-mediated transcriptional control has been used to mediate differentiation of cells towards neural lineages [521], endodermal lineages [558], and adipocyte-like cells [559], transdifferentiation of mouse embryonic fibroblasts towards myocytes [560], and neurons [561], and conversion of mouse hepatic cells towards pancreas by Pdx1 activation *in vivo* [562]. In pluripotent reprogramming context, CRISPR-mediated gene activation has been used for activation of pluripotent reprogramming factors in human fibroblasts [563], conversion of human PSCs to naïve state by *NANOG* activation [525], and replacement of transgenic Oct4 in mouse pluripotent reprogramming by enhancer targeted dCas9 activators [564]. The reprogramming of mouse fibroblasts to pluripotency was also recently demonstrated by endogenous *Oct4* and *Sox2* activation by CRISPRa [565]. The mouse reprogramming process was shown to be improved by a cocktail of small molecular compounds, demonstrating the potential in combining targeted epigenome editing with additional reprogramming enhancing approaches. However, replication of transgenic transcription factor mediated cell type conversions with CRISPR-mediated gene activation have so far only been limited, and the methods thereof need to be further developed for robust CRISPR-mediated cell type conversions from variable starting cell sources.

6.5. Transposable Elements in Human Genome

A major portion of human genome consists of transposable elements (TE). These are generally divided into DNA transposons, LTR retrotransposons, or endogenous retroviruses, and non-LTR retrotransposons including long interspersed nuclear elements (LINE) and short interspersed nuclear elements (SINE) [566]. Majority of human TEs have lost their capacity to actively transpose, but some elements, such as some HERV-K and LINE-1 copies, may still actively move [566]. Therefore, it is important for a cell to control the transposition of mobile elements, to limit the insertional mutagenesis caused by TE integrations. Cells thus employ a large set of KRAB zinc-finger proteins and other factors to target endogenous TEs and to control their expression [347,567]. In addition to direct insertional mutations, integration of TEs near genes may affect the expression of the associated genes and result in rewiring of gene regulatory networks, for example by providing new enhancer elements, alternative promoters or mRNA splice sites. Integration of new TEs and exaptation of ancestral TEs by point mutations can contribute to evolutionary adaptation and the species specific differences in transcriptional networks [568]. It has been reported that only a minority of OCT4 and NANOG binding sites occur in homologous sites between mouse and human pluripotent stem cells and that TEs contribute to a large fraction of the binding sites of these factors [569].

6.5.1. Expression of Transposable Elements in Development, Pluripotency and Reprogramming

Early development mammalian embryos express distinct classes of transposable elements. Some of the proteins important for the early embryo development, such as the trophoblast expressed Syncytin, are originally of viral origin [570], and it has been reported that human embryos can even produce virus-like particles from endogenous retroviral sequences [571]. The role of transposable element expression in early embryo development is not fully clear, but mouse embryo experiments suggest a role for some TEs in promoting global chromatin opening in embryonic genome activation [572,573]. The expression of transposable elements in embryo development is dynamically regulated and early embryonic cells show distinct cell type specific and temporal patterns of transposable element expression [574]. The dynamic pattern of TE expression in embryos can be used to distinguish various pluripotent states in cell culture, likely reflecting the differences in expression of the factors targeting the TEs. For example, the blastocyst expressed HERVH derived LTR7 promoter has been used to distinguish naïve-like pluripotent cell populations [198,574], although contradicting results have later been reported [575], possibly reflecting differences in the various human naïve states. In mouse pluripotent stem cells, the early embryo expressed MERVL LTR, which is activated by DUX protein, has been used to mark rare sub populations of cultured cells that are capable of chimeric contribution into both embryonic and extraembryonic tissues [14,576]. Overall, due to the high number and differential expression

of transposable elements in pluripotent cells, their expression pattern has been suggested as a method to characterise alternative pluripotent states [575].

In reprogramming, transposable elements have been reported to lose their repressed status resulting in transient activation of TE expression [577]. This appears to also affect the expression of TE associated genes, indicative of promoter or enhancer-like effect of the elements in the process [577]. Many of the transiently activated TEs are later re-silenced in the fully reprogrammed pluripotent cell populations, possibly due to re-activation of the TE targeting factors, such as the KRAB-ZFPs. The role of the TE activation in the reprogramming process is not fully understood, however, it has been reported that knock down of the HERV-H LTR7 transcript negatively affects reprogramming and greatly reduces reprogramming efficiency [199]. The same LTR7 expression, along with KLF4, has also been implicated in the differentiation defective phenotype of human iPSCs, and therefore the proper silencing of these elements may be required for stable pluripotency [196]. Additionally, retroelement re-activation during reprogramming has been reported to result in retrotransposition of endogenous Alu- and L1-elements [578]. The activation of transposable elements during the reprogramming process may thus additionally contribute to mutations occurring in reprogramming.

As our ability to target genomic elements in a sequence specific manner has been rapidly increasing following the developments in the CRISPR field, we will be able to study the contribution of the endogenous transposable elements to various cellular processes in more detail.

6.5.2. Alu Elements

Alu elements are among the most abundant human SINE sequences, with more than one million copies in the human genome. Alus were initially described as AluI restriction enzyme recognition site containing repeat sequences in renatured human DNA [579]. Alu monomers, and the related B1 sequence precursors in mouse, were originally derived from the 7SL RNA by deletion of a central 7SL-specific sequence approximately 90 million years ago [580,581]. The full length Alu sequence consists of two similar but distinct Alu monomers that have been fused in a head-to-tail orientation, resulting in an approximately 300 bp transposable element sequence [582]. The origin and amplification of Alu sequences happened within the past 65 million years in primates with majority of the Alu amplification occurring over 40 million years ago [583].

Alu elements have been implicated in various forms of gene expression regulation, including alternative splicing, RNA editing and translational control [584]. Alu genomic distribution is biased and Alu sequences are found enriched in genome topological boundary regions and near upstream regions and downstream intronic regions of specific classes of genes [585,586]. Functionally some Alu and B1 elements have been implicated in genomic insulation [587,588]. Particularly, a specific set of B1 elements have been reported to exhibit an AHR and SLUG mediated transcriptional activation dependent chromatin insulation by recruitment of PARP1 and CTCF

to the transcribed element [588]. Alu sequences have also been described to have proto-enhance-like histone modification patterns and preferential interactions with gene transcription start sites, suggesting a direct role for Alu elements in transcriptional control [589]. This is also implicated by TFIIIC and Alu sequence mediated recruitment of activated genes to transcription factories, and Pol III transcribed Alus having active enhancer function in specific neural genes [590,591]. Alu elements also contain many binding sites for various transcription factors [592,593]. Particularly, Alu sequences have been reported to bind YY1, which may link Alus to chromatin architectural features required for gene transcription control [594].

Alu sequences have been reported to be enriched near the first genes expressed at human embryo genome activation and near pluripotent state specific genes [595,596]. Alu elements also contain binding sites for many early embryo expressed factors, like totipotent cell PRD-like factors [597,598], HNF4α [599], AHR [600], and retinoic acid receptors [601]. It is therefore possible, that Alu sequences may function in controlling the acquisition of pluripotent state or differentiation. However, the impact of these elements on pluripotent reprogramming is currently unknown. Schematic representation of Alu element structure and selected transcription factor binding sites are presented in **Figure 11**.

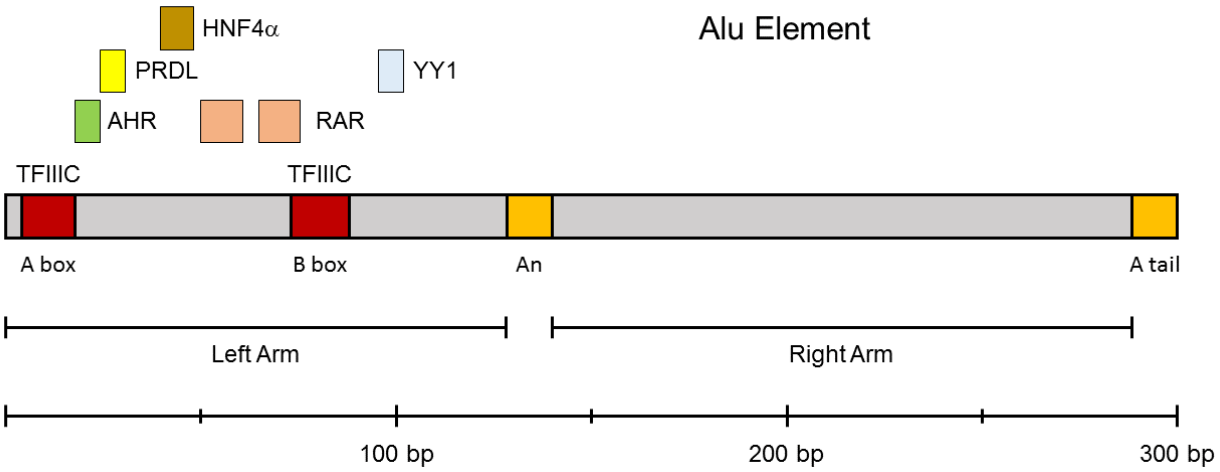

Figure 11. *Structure of Alu elements. Alu element consensus sequences contain two head-to-tail fused monomers. Left Alu arm contains Pol III promoter core motifs and binding sites for many early embryo expressed factors. A and B boxes are Pol III promoter core domains binding TFIIIC. TFIIIC = Transcription factor for polymerase III C, AHR = Aryl hydrocarbon receptor, RAR = Retinoic acid receptor, PRDL = PAIRED (PRD)-like homeobox, HNF4α = Hepatocyte nuclear factor 4 alpha, YY1 = Ying Yang 1, An = mid A-stretch, A tail = terminal A-stretch.*

7. Aims of the Study

Induction of pluripotency has been intensively studied during the past decade since its discovery. Although the process is generally robust, it can often result in suboptimal reprogramming outcomes. The general aim of this thesis was to investigate factors limiting pluripotent reprogramming and to develop new approaches to improve the outcome of the reprogramming process.

The specific aims of this thesis were:

1) To investigate the use of AAV as an episomal transgene delivery vector to produce integration-free induced pluripotent stem cells.

2) To investigate the impact of age and passaging on pluripotent reprogramming efficiency.

3) To develop the CRISPR gene activation system for targeting multiple endogenous genes as a tool for reprogramming somatic cells.

8. Materials and Methods

8.1. Ethics Statement

The generation and use of human induced pluripotent stem cell lines used in this study was approved by the Coordinating Ethics Committee of the Helsinki and Uusimaa Hospital District (nr. 423/13/03/00/08).

8.2. Cell Culture

Human skin fibroblasts, mouse embryonic fibroblasts, HEK239 and HEK239T cells were cultured in fibroblast medium consisting of Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 2 mM GlutaMAX, and 100 µl/ml penicillin-streptomycin.

HEK293GPG cells, used for VSV-G pseudotyped γ-retrovirus production, were cultured in fibroblast medium supplemented with 2 mg/ml Puromycin, 6 mg/ml G418 and 1 µg/ml Tetracycline to repress viral protein expression. For retrovirus production, antibiotics were removed from the culture medium.

Mouse iPSCs (**I**) were cultured on mitomycin-C treated mouse embryonic fibroblast (MEF) feeders in mES medium consisting of KnockOut high-glucose DMEM (at 4,500 mg/l) supplemented with 1 mM sodium pyruvate, 15% FBS, 2 mM GlutaMAX, 0.10 mM non-essential amino acids (NEAA), 0.1 mM β-mercaptoethanol, and leukemia inhibitory factor (LIF). Mouse iPSC were split using TrypLE Select.

Human iPSCs (**I**, **II**, **III** and **IV**) were cultured on Matrigel coated cell culture plates in either MEF-conditioned hES medium consisting of KnockOut DMEM supplemented with 20% KO serum replacement, 2mM GlutaMAX, 0.1 mM β-mercaptoethanol, 1% NEAA, and 6 ng/ml basic fibroblast growth factor (bFGF) (**I**), or on mitomycin-C treated MEF feeders in hES medium (**I**) or on Matrigel coated plates in E8 medium (**II**, **III**, **IV**). Cells cultured in hES medium (**I**) were split by combination of collagenase IV and mechanic dissociation, and cell cultured in E8 medium (**II** ,**III**, **IV**) were split by dissociation with 0.5 mM EDTA.

All cell lines were cultured in an incubator at +37°C and 5% CO_2.

8.3. AAV Reprogramming (I)

Recombinant Adeno-associated viruses were produced at the Biomedicum AAV Gene Transfer and Cell Therapy Core Facility as described by Zolotukhin *et al*. [602]. Briefly, three 15 cm dishes per factor of HEK293T cells were transfected with three parts pDG plasmid, encoding AAV2 capsid proteins, and one part pSub-CMV-WPRE or pSubCAG-WPRE plasmid containing the reprogramming transgenes. 34.5 µg of plasmid was used per plate and transfection was performed using JetPEI transfection reagent. Two days later cells were collected and lysed to release virus particles by three freeze-thaw cycles, DNAse treatment and

centrifugation. Viruses were purified from supernatant by iodixanol gradient ultracentrifugation, and virus amounts were quantified by qPCR.

For human foreskin fibroblast (HFF; ATCC, CRL-2429) reprogramming attempts with rAAV2, cells were seeded on 6-well plates in fibroblast medium, 100 000 cells per well, and infected with Yamanaka factor encoding rAAV2 viruses 1 to 4 times every second day. On day four cells were split on mitomycin-C treated MEF feeders or Matrigel coated wells and medium was changed to hES medium. Cell culture medium was changed every second day.

For mouse embryonic fibroblast reprogramming with rAAV2, cells were seeded on 6-well plates in fibroblast medium, 100 000 cells per well, and infected with Yamanaka factor encoding rAAV2 viruses 1 to 3 times every second day with multiplicity of infection (MOI) between 20 000 and 200 000 viral genomes per cell per factor. On day four cells were split on gelatin coated cell culture plates and medium was changed to mES medium. Emerging colonies were picked manually onto mitomycin-C treated MEF feeders in mES medium.

8.4. Retroviral Reprogramming (II)

VSV-G pseudotyped γ-retroviruses were produced by transfecting approximately 90% confluent 6-well plate wells of HEK293GPG cells with 3µg of pMXs plasmids encoding Yamanaka factors using FuGENE HD reagent. Cells were transfected twice to ensure robust transfection. After second transfection cell culture medium was changed to remove antibiotics and allow virus production. Virus containing media were collected 4, 5 and 6 days after first transfection, pooled and filtered through 0.45 µm syringe filters. Fibroblasts were seeded on 6-well plates at 100 000 cells per well on day 4 of virus production. Fibroblasts were transduced with the viruses on days 5 and 6 overnight, with 1:1 mixture of virus medium and fresh fibroblast medium. Cells were split 3 days after the first virus transduction onto Matrigel coated plates and cell culture medium was changed to hES medium or E6 medium supplemented with 0.25 mM sodium butyrate. Emerging colonies were manually picked on Matrigel coated cell culture dishes in E8 medium (Life Technologies).

8.5. Sendai Viral Reprogramming (II)

Sendai viral reprogramming was performed using CytoTune Sendai viral reprogramming kit (Life Technologies). Briefly, 100 000 fibroblasts were infected with a MOI of 3 in fibroblast medium. Five to seven days later cells were split on Matrigel coated dishes and cell culture medium was changed to E6 medium (Life Technologies) supplemented with 0.25 mM sodium butyrate. Emerging colonies were manually picked on Matrigel coated cell culture dishes in E8 medium (Life Technologies).

8.6. Episomal CRISPRa Reprogramming (III, IV)

For episomal plasmid reprogramming human fibroblasts were dissociated with TrypLE into single cell solution and eletroporated with Neon transfection system (Thermo Fisher) using 100µl tips, 1 million cells, 6 µg total DNA amount and 1650 V, 10 ms and 3 x pulse settings. For OCT4 replacement with dCas9 activator (**III**), 1.5 µg of each pCXLE-dCas9VP192-GFP-shP53 (Addgene plasmid #69535), GG-EBNA-OCT4 (Addgene plasmid #69537), pCXLE-hSK (Addgene plasmid #27078) and pCXLE-hUL (Addgene plasmid #27080) were used. For full CRISPRa reprogramming with episomal plasmids (**IV**), 2 µg of dCas9 activator plasmid and 4 µg of guide RNA plasmids were used. Electroporated cells were plated on gelatin coated plates and split onto Matrigel coated 10 cm dishes on day six (**III**) or directly plated on Matrigel coated dishes in fibroblast medium (**IV**). After four days cell culture medium was changed to mixture of fibroblast medium and hES medium supplemented with 0.25 mM sodium butyrate. Emerging colonies were manually picked on Matrigel coated 24-well plate wells in E8 medium.

8.7. Transposon CRISPRa Reprogramming (IV)

For transposon CRISPRa reprogramming human fibroblasts were electroporated as described for episomal reprogramming, but with PiggyBac transposon plasmids and 3.5 µg total DNA amount. Plasmids used were 1 µg PB-tight-DDdCas9VP192-GFP-IRES-Neo, 1 µg PB-CAG-rtTAM2-IN, 1 µg PB-EEA-5g-$OSK_2M_2L_1$-PGK-Puro and 0.5 µg PiggyBac transposase plasmid per electroporation. Electroporated cells were plated on gelatin coated cell culture dishes in fibroblast medium. Four days after electroporation cells were selected with 1 µg/ml Puromycin and 0.5 mg/ml G418 for two days after which antibiotic concentrations were halved. To induce reprogramming, cells were plated on Matrigel coated dishes in 1:1 mixture of fibroblast medium and hES medium supplemented with 0.25 mM sodium butyrate. To induce dCas9 activator expression cell culture medium was supplemented with 1 µM trimethoprim (TMP) and 2 µg/ml doxycycline (DOX). Fresh DOX was supplemented daily. Emerging colonies were manually picked on Matrigel coated plates in E8 medium without TMP and DOX.

8.8. Alkaline Phosphatase Staining (I, II, III, IV)

Cells were fixed with 4% paraformaldehyde (PFA) for 10 minutes after which cells were washed with phosphate buffered saline. After fixation cells were stained with NBT/BCIP solution in 0.1 M Tris HCl pH 9.5, 0.1 M NaCl, 0.05 M $MgCl_2$ buffer until purple precipitate formed. Reaction was stopped by washing the plates and colonies were counted based on alkaline phosphatase positive staining and compact iPSC-like morphology.

8.9. Immunocytochemistry

For immunocytochemical staining cells were fixed wit 4 % PFA for 30 minutes, permeabilised with 0.2 % Triton X-100 in PBS for 15 minutes and blocked with Ultra V Block (Thermo Fisher) for 10 minutes. Thereafter cells were incubated in primary antibody containing PBS with 0.1 % Tween 20 overnight in +6 °C with indicated primary antibody amounts or for 2 days with halved antibody amounts. Secondary antibody staining was done in PBS with 0.1 % Tween 20 for half an hour in room temperature in the presence of Hoechst33342.

Table 1. *Antibodies used in this work*

Antibody	dilution	source	identifier	article	use
LIN28A	1:250	Cell Signaling	D84C11	**III, IV**	ICC
	1:250	Cell Signaling	D1A1A	**IV**	ICC
	1:100	Thermo Fisher	MA1-016	**III**	ICC
NANOG	1:250	Cell Signaling	D73G4	**IV**	ICC
	1:100	Thermo Fisher	MA1-017	**III**	ICC
	1:100	Santa Cruz	sc-30331	**I**	ICC
OCT4	1:500	Santa Cruz	sc-8628	**I, III, IV**	ICC
	1:500	Cell Signaling	C30A3	**II, III**	ICC
	1:100	Thermo Fisher	MA1-104	**III**	ICC
	1:1000	Santa Cruz	sc-9081	**I**	WB
SOX2	1:250	Cell Signaling	D6D9	**III, IV**	ICC
	1:200	Millipore	ab5603	**I**	ICC
	1:2500	Millipore	ab5603	**I**	WB
KLF4	1:250	Sigma-Aldrich	HPA002926	**III, IV**	ICC
	1:00	Abcam	ab34814	**I**	ICC
	1:00	Abcam	ab34814	**I**	WB
C-MYC	1:250	Cell Signaling	D3N8F	**IV**	ICC
	1:250	Abcam	[Y69]ab32072	**IV**	ICC
	1:25	Santa Cruz	(c-33) sc-42	**I**	ICC
TRA-1-60	1:50	Thermo Fisher	MA1-023	**II, IV**	ICC, live
TRA-1-81	1:100	Thermo Fisher	MA1-024	**IV**	ICC
TUBB3	1:500	R&D Systems	MAB1195	**IV**	ICC
AFP	1:400	Dako	A0008	**IV**	ICC
SMA	1:200	Sigma-Aldrich	A2547	**IV**	ICC
VIMENTIN	1:500	Santa Cruz	sc-5565	**IV**	ICC
SOX17	1:500	R&D Systems	AF1924	**III, IV**	ICC
CDH1	1:500	BD Bioscience	610181	**I, III**	ICC
FOXA2	1:500	Santa Cruz	sc-9187	**III**	ICC
PDX1	1:200	R&D Systems	AF2419	**III**	ICC
NKX6.1	1:200	Developmental studies hybridoma bank	F55A10	**III**	ICC

GATA4	1:500	Santa Cruz	sc-1237	**III**	ICC
MAFA	1:500	Abcam	AB26405	**III**	ICC
SSEA3	1:100	Millipore	MAB4303	**II**	ICC
SSEA1	1:50	Millipore	MAB4301	**I**	ICC
DESMIN	1:50	Santa Cruz	sc-14026	**I**	ICC
donkey anti-goat	1:500	Invitrogen	A11055	**IV**	2nd
	1:500	Invitrogen	A11058	**IV**	2nd
donkey anti-mouse	1:500	Invitrogen	A21202	**IV**	2nd
	1:500	Invitrogen	A21203	**IV**	2nd
donkey anti-rabbit	1:500	Invitrogen	A21206	**IV**	2nd
	1:500	Invitrogen	A21207	**IV**	2nd

ICC: immunocytochemistry, WB: western blot, live: live staining, 2nd; secondary antibody

8.10. Reverse Transcription Polymerase Chain Reaction

Total RNA was extracted using NucleoSpin RNA kit (Macherey-Nagel) according to manufacturer's instructions. One or two micrograms of denatured RNA were used for cDNA synthesis using 0.5 µL Moloney murine leukemia virus (MMLV) reverse transcriptase, 0.2 µL Random Primers, 1 µL Oligo(dT)18 Primer and 0.5 µL Ribolock RNase inhibitor for 90 min at 37°C. Mouse pluripotency factor RT-PCR (**I**) was run for 35 cycles with 60°C annealing temperature. Quantitative RT-PCR was performed as described by Toivonen *et al*. [603]. Human pluripotency factor quantitative RT-PCR (**III**, **IV**) were run for 40 cycles with 57°C annealing temperature 25 sec extension time and human fibroblast qRT-PCR (**II**) were run with 60°C annealing temperature 30 sec extension time.

Table 2. *RT-PCR primers used in this work*

Factor	**forward**	**reverse**	**species**	**article**
Oct4	TTGGGCTAGAGAAGGATGTG	GAGTAGAGTGTGGTGAAGTG	mouse	**I**
Sox2	GAAACGACAGCTGCGGAAA	TCTAGTCGGCATCACGGTTTT	mouse	**I**
Klf4	AACATGCCCGGACTTACAAA	TTCAAGGGAATCCTGGTCTTC	mouse	**I**
Myc	TAACTCGAGGAGGAGCTGGA	GCCAAGGTTGTGAGGTTAGG	mouse	**I**
Nanog	CTCAGCCTCCAGCAGATGC	GGGATAGCTGCAATGGATGC	mouse	**I**
Rex1	TGTCCTCAGGCTGGGTAGTC	TGATTTTCCGACGTATGCAA	mouse	**I**
OCT4 (tg)	TCCCCCTGTCTCCGTCACCAC	GCGGCCCAAAGGGAGATCCG	transgene	**I**
SOX2 (tg)	CCTACTCGCAGCAGGGCACC	GCGGCCCAAAGGGAGATCCG	transgene	**I**
KLF4 (tg)	TAAACACACGGGGCACCGCC	GCGGCCCAAAGGGAGATCCG	transgene	**I**
Myc (tg)	AAACGACAAGAGGCGGACAC	GCGGCCCAAAGGGAGATCCG	transgene	**I**
WPRE	TGAGTTTGGACAAACCACAAC	TTGTTGTTAACTTGTTTATTGCAGC	transgene	**I**
p16	GTGAGAGTGGCGGGGTC	TCGTGCTGATGCTACTGAGG	human	**II**
p21	AGCTGCTCGCTGTCCACT	CTGCGTTCACAGGTGTTTCT	human	**II**
p53	TGTTTCCTGACTCAGAGGGG	GAGCGTGCTTTCCACGAC	human	**II**
PUMA	CACCTAATTGGGCTCCATCT	GACCTCAACGCACAGTACGA	human	**II**

OCT4	TTGGGCTCGAGAAGGATGTG	TCCTCTCGTTGTGCATAGTCG	human	**II, III, IV**
SOX2	GCCCTGCAGTACAACTCCAT	TGCCCTGCTGCGAGTAGGA	human	**II, III, IV**
NANOG	CTCAGCCTCCAGCAGATGC	TAGATTTCATTCTCTGGTTCTGG	human	**II, III, IV**
TDGF1	TCAGAGATGACAGCATTTGGC	TTCAGGCAGCAGGTTCTGTTTA	human	**II**
SeV	GGATCACTAGGTGATATCGAGC	ACCAGACAAGAGTTTAAGAGATATGTATC	transgene	**II**
GAPDH	GGTCATCCATGACAACTTTGG	TGAGCTTCCCGTTCAGCTC	human	**II**
CYCLOG	TCTTGTCAATGGCCAACAGAG	GCCCATCTAAATGAGGAGTTG	human	**III, IV**
LIN28	AGGAGACAGGTGCTACAACTG	TCTTGGGCTGGGGTGGCAG	human	**III, IV**
KLF4	CCGCTCCATTACCAAG	CACGATCGTCTTCCCCTCTT	human	**III, IV**
CDH1	ATGAGTGTCCCCCGGTATCT	GGTCAGTATCAGCCGCTTTC	human	**III, IV**
FOXA2	AAGACCTACAGGCGCAGCT	CATCTTGTTGGGGCTCTGC	human	**III, IV**
SOX17	CCGAGTTGAGCAAGATGCTG	TGCATGTGCTGCACGCGCA	human	**III**
GATA4	GAGGAAGGAGCCAGCCTAGCAG	CGGGTCCCCCACTCGTCA	human	**III, IV**
PDX1	AAGTCTACCAAAGCTCACGCG	CGTAGGCGCCGCCTGC	human	**III**
NKX6.1	TATTCGTTGGGGATGACAGAG	TGGCCATCTCGGCAGCGTG	human	**III**
NKX2.2	GAACCCCTTCTACGACAGCA	ACCGTGCAGGGAGTACTGAA	human	**III**
SOX9	ATCAAGACGGAGCAGCTGAG	GGCTGTAGTGTGGGAGGTTG	human	**III**
MAFA	GCCAGGTGGAGCAGCTGAA	CTTCTCGTATTTCTCCTTGTAC	human	**III**
GCK	CCGCCAAGAAGGAGAAGGTA	CTTCTGCATCCGTCTCATCA	human	**III**
MYC	AGCGACTCTGAGGAGGAACA	CTCTGACCTTTTGCCAGGAG	human	**IV**
REX1	CGTTTCGTGTGTCCCTTTCAA	CCTCTTGTTCATTCTTGTTCGT	human	**IV**
ASCL1	ACTCGTCGGACGAGGGCTCTTA	GCACTAAAGATGCAGGTTGTGCGA	human	**IV**
NEUROG2	ATCCGAGCAGCACTAACACG	GCACAGGCCAAAGTCACAG	human	**IV**
dCas9	AAACAGCAGATTCGCCTGGA	TCATCCGCTCGATGAAGCTC	transgene	**IV**
mCherry	CCACTACGACGCTGAGGTCAA	TCGTTGTGGGAGGTGATGTCC	transgene	**IV**

8.11. Embryoid Body Assay

For mouse embryoid body (EB) assay (**I**), iPSCs were dissociated with TrypLE and plated on cell culture dishes for 30 mins to allow feeders to attach. Thereafter iPSCs were split onto non-adherent cell culture dishes in mES medium without LIF. Embryoid bodies were grown in suspension for 14 days after which cells were fixed and stained.

For human embryoid body assay (**III**, **IV**), iPSCs were dissociated with EDTA and plated on non-adherent cell culture dishes in hES medium without bFGF supplemented with 5µM ROCK inhibitor (Y27632). The next day medium was changed to hES without bFGF and without ROCK inhibitor. Embryoid bodies were grown for two weeks after which cells were plated on gelatin coated cell culture dishes for a week to allow outgrowth. Thereafter cells were fixed and stained.

8.12. Guide RNA Design and Production (III, IV)

Guide RNA sequences were designed using the Zhang lab CRISPR guide design tool (http://crispr.mit.edu/). Guides were targeted to sites -400 bp to -50 bp upstream of the transcription start site and selected for further testing based on their position and off-target score. Guide RNA expression cassettes were assembled in a PCR reaction containing PCR amplified U6 promoter template, tracrRNA and terminator containing template and a single stranded guide oligo containing 20 bp overlap with the flanking fragments. PCR reaction conditions contained 50 pmol forward and reverse primers, 2 pmol guide oligo, 5 ng U6 and tracrRNA fragments and were amplified with Phusion polymerase (Thermo Fisher) in 100 micro litre reaction volume. PCR reaction program was 98°C, 10 sec; 52°C, 30 sec; 72°C, 12 sec for 35 cycles. Successful PCR assembly of guide cassettes was checked with gel electrophoresis and guides were purified with a column and confirmed with sequencing.

Guide RNA assembly reaction:

5x HF buffer	20 µl
dNTP (2.5 mM)	8 µl
gRNA (1uM)	2 µl
1aggc Fw (100 uM)	0,5 µl
1aggc Rv (100 uM)	0,5 µl
Phusion (Thermo Scientific)	1 µl
U6 prom Tailed (5 ng)	x
Term tailed (5 ng)	x
H_2O	x (up to 100 µl)
Total	100 µl

Thermal cycle:

1x	98°C - 3’
35x	98°C - 10’’
	52°C - 30’’
	72°C - 12’’
1x	72°C - 8’

Table 3. *Guide oligo sequences used in this work*

target	guide nr	sequence	article
OCT4	1	GTGGAAAGGACGAAACACCG GGGGGAGAAACTGAGGCGA GTTTTAGAGCTAGAAATAG	**III, IV**
OCT4	2	GTGGAAAGGACGAAACACCG GGTGGTGGCAATGGTGTCTG GTTTTAGAGCTAGAAATAG	**III, IV**
OCT4	3	GTGGAAAGGACGAAACACCG GACACAACTGGCGCCCCTCC GTTTTAGAGCTAGAAATAG	**III, IV**
OCT4	4	GTGGAAAGGACGAAACACCG GGCACAGTGCCAGAGGTCTG GTTTTAGAGCTAGAAATAG	**III, IV**
OCT4	5	GTGGAAAGGACGAAACACCG TCTGTGGGGGACCTGCACTG GTTTTAGAGCTAGAAATAG	**III, IV**
SOX2	1	GTGGAAAGGACGAAACACCG TGTAAGGTAAGAGAGGAGAG GTTTTAGAGCTAGAAATAG	**III, IV**
SOX2	2	GTGGAAAGGACGAAACACCG TTTACCCACTTCCTTCGAAA GTTTTAGAGCTAGAAATAG	**III, IV**
SOX2	3	GTGGAAAGGACGAAACACCG GTGGCTGGCAGGCTGGCTCT GTTTTAGAGCTAGAAATAG	**III, IV**
SOX2	4	GTGGAAAGGACGAAACACCG CAAAACCCGGCAGCGAGGCT GTTTTAGAGCTAGAAATAG	**III, IV**
SOX2	5	GTGGAAAGGACGAAACACCG AGGAGCCGCCGCGCGCTGAT GTTTTAGAGCTAGAAATAG	**III, IV**
KLF4	1	GTGGAAAGGACGAAACACCG CGAACGTGTCTGCGGGCGCG GTTTTAGAGCTAGAAATAG	**III, IV**

KLF4	2	GTGGAAAGGACGAAACACCG TATAAGTAAGGAACGCGCGC GTTTTAGAGCTAGAAATAG	**III, IV**
KLF4	3	GTGGAAAGGACGAAACACCG GCTGCCATAGCAACGATGGA GTTTTAGAGCTAGAAATAG	**III, IV**
KLF4	4	GTGGAAAGGACGAAACACCG GTTCGGTCGCTGCGCGACCA GTTTTAGAGCTAGAAATAG	**III, IV**
KLF4	5	GTGGAAAGGACGAAACACCG TCTTCGCGGGCTTCGAACCC GTTTTAGAGCTAGAAATAG	**III, IV**
MYC	1	GTGGAAAGGACGAAACACCG CCCTTTATAATGCGAGGGTC GTTTTAGAGCTAGAAATAG	**IV**
MYC	2	GTGGAAAGGACGAAACACCG TCTCGCTAATCTCCGCCCAC GTTTTAGAGCTAGAAATAG	**IV**
MYC	3	GTGGAAAGGACGAAACACCG GGTTCCCAAAGCAGAGGGCG GTTTTAGAGCTAGAAATAG	**IV**
MYC	4	GTGGAAAGGACGAAACACCG AGCTAGAGTGCTCGGCTGCC GTTTTAGAGCTAGAAATAG	**IV**
MYC	5	GTGGAAAGGACGAAACACCG GCGCGCGTAGTTAATTCATG GTTTTAGAGCTAGAAATAG	**IV**
LIN28A	1	GTGGAAAGGACGAAACACCG GTGTCAGAGACCGGAGTTGT GTTTTAGAGCTAGAAATAG	**III, IV**
LIN28A	2	GTGGAAAGGACGAAACACCG CCCATCTCCAGTTGTGCGTG GTTTTAGAGCTAGAAATAG	**III, IV**
LIN28A	3	GTGGAAAGGACGAAACACCG CGGGGTACTCAAGTCTTCTA GTTTTAGAGCTAGAAATAG	**III, IV**
LIN28A	4	GTGGAAAGGACGAAACACCG TAATTATCTGCCCGGGGGGT GTTTTAGAGCTAGAAATAG	**III, IV**
LIN28A	5	GTGGAAAGGACGAAACACCG TCTGATTGGCCAGCGCCGCC GTTTTAGAGCTAGAAATAG	**III, IV**
NANOG	1	GTGGAAAGGACGAAACACCG TCCCAATTTACTGGGATTAC GTTTTAGAGCTAGAAATAG	**III, IV**
NANOG	2	GTGGAAAGGACGAAACACCG TGATTTAAAAGTTGGAAACG GTTTTAGAGCTAGAAATAG	**III, IV**
NANOG	3	GTGGAAAGGACGAAACACCG TCTAGTTCCCCACCTAGTCT GTTTTAGAGCTAGAAATAG	**III, IV**
NANOG	4	GTGGAAAGGACGAAACACCG GATTAACTGAGAATTCACAA GTTTTAGAGCTAGAAATAG	**III, IV**
NANOG	5	GTGGAAAGGACGAAACACCG CGCCAGGAGGGGTGGGTCTA GTTTTAGAGCTAGAAATAG	**III, IV**
EEA	1	GTGGAAAGGACGAAACACCG CCCAGCACTTTGGG GTTTTAGAGCTAGAAATAG	**IV**
EEA	2	GTGGAAAGGACGAAACACCG AATCCCAGCACTTT GTTTTAGAGCTAGAAATAG	**IV**
EEA	3	GTGGAAAGGACGAAACACCG GCCTCCCAAAGTGC GTTTTAGAGCTAGAAATAG	**IV**
EEA	7	GTGGAAAGGACGAAACACCG GCTACTTGGGAGGC GTTTTAGAGCTAGAAATAG	**IV**
EEA	10	GTGGAAAGGACGAAACACCG GCCTCCCAAGTAGC GTTTTAGAGCTAGAAATAG	**IV**
TdT	1	GTGGAAAGGACGAAACACCG GAGTTCGAGATCGA GTTTTAGAGCTAGAAATAG	**IV**
TdT	2	GTGGAAAGGACGAAACACCG TTACGGGGCCGTCG GTTTTAGAGCTAGAAATAG	**IV**
TdT	3	GTGGAAAGGACGAAACACCG AGCACGCCGTCGCG GTTTTAGAGCTAGAAATAG	**IV**
TdT	4	GTGGAAAGGACGAAACACCG GGCCGCCCCTACGA GTTTTAGAGCTAGAAATAG	**IV**
TdT	5	GTGGAAAGGACGAAACACCG CGTGATGAACTTCG GTTTTAGAGCTAGAAATAG	**IV**
common ctrl	7	GTGGAAAGGACGAAACACCG GATTTTTAGTAGAGA GTTTTAGAGCTAGAAATAG	**IV**
common ctrl	8	GTGGAAAGGACGAAACACCG GTGGGAGGCTGAGGC GTTTTAGAGCTAGAAATAG	**IV**
common ctrl	9	GTGGAAAGGACGAAACACCG GAGTGCTGGGATTAC GTTTTAGAGCTAGAAATAG	**IV**
common ctrl	10	GTGGAAAGGACGAAACACCG GGTAGCTGGGATTAC GTTTTAGAGCTAGAAATAG	**IV**
common ctrl	11	GTGGAAAGGACGAAACACCG GCATGTTGGCCAGGC GTTTTAGAGCTAGAAATAG	**IV**
ASCL1	1	GTGGAAAGGACGAAACACCG CGGGAGAAAGGAACGGGAGG GTTTTAGAGCTAGAAATAG	**IV**
ASCL1	2	GTGGAAAGGACGAAACACCG AAGAACTTGAAGCAAAGCGC GTTTTAGAGCTAGAAATAG	**IV**
ASCL1	3	GTGGAAAGGACGAAACACCG TCCAATTTCTAGGGTCACCG GTTTTAGAGCTAGAAATAG	**IV**
ASCL1	4	GTGGAAAGGACGAAACACCG GTTGTGAGCCGTCCTGTAGG GTTTTAGAGCTAGAAATAG	**IV**
NGN2	1	GTGGAAAGGACGAAACACCG GGCGGTGGCGGGGGAGGAGG GTTTTAGAGCTAGAAATAG	**IV**
NGN2	2	GTGGAAAGGACGAAACACCG CAATGAAAAGAATAAGCCAG GTTTTAGAGCTAGAAATAG	**IV**
NGN2	3	GTGGAAAGGACGAAACACCG GGGAAAGGCGGTGAAGAAAG GTTTTAGAGCTAGAAATAG	**IV**
NGN2	4	GTGGAAAGGACGAAACACCG CGGAGCTGGCGAAGCCGCAG GTTTTAGAGCTAGAAATAG	**IV**
GATA4	1	GTGGAAAGGACGAAACACCG ACCTCCAAGGAATCCGGGGC GTTTTAGAGCTAGAAATAG	**III, IV**
GATA4	2	GTGGAAAGGACGAAACACCG CTCAACTCTCGATCTTGTGT GTTTTAGAGCTAGAAATAG	**III, IV**

GATA4	3	GTGGAAAGGACGAAACACCG	CAGCGAACCCAATCGACCTC	GTTTTAGAGCTAGAAATAG	**III, IV**
GATA4	4	GTGGAAAGGACGAAACACCG	AATGCCCAAGTGCTACCGCC	GTTTTAGAGCTAGAAATAG	**III, IV**
GATA4	5	GTGGAAAGGACGAAACACCG	CCTGTGGGAGTCACGTGCAA	GTTTTAGAGCTAGAAATAG	**III, IV**
FOXA2	1	GTGGAAAGGACGAAACACCG	AGTGCCGAGCTGCCCCGAGG	GTTTTAGAGCTAGAAATAG	**III, IV**
FOXA2	2	GTGGAAAGGACGAAACACCG	CGCGCGGCGCGGGGGCTAGT	GTTTTAGAGCTAGAAATAG	**III, IV**
FOXA2	3	GTGGAAAGGACGAAACACCG	TGCGGCACTTGTCCGCTCCG	GTTTTAGAGCTAGAAATAG	**III, IV**
FOXA2	4	GTGGAAAGGACGAAACACCG	TATAGCGCGGCGCGCTGGCG	GTTTTAGAGCTAGAAATAG	**III, IV**
FOXA2	5	GTGGAAAGGACGAAACACCG	AAATGGGCTGCCCCGGGTCT	GTTTTAGAGCTAGAAATAG	**III, IV**
CDH1	1	GTGGAAAGGACGAAACACCG	AGGGTCACCGCGTCTATGCG	GTTTTAGAGCTAGAAATAG	**III, IV**
CDH1	2	GTGGAAAGGACGAAACACCG	CAGTGGAATCAGAACCGTGC	GTTTTAGAGCTAGAAATAG	**III, IV**
CDH1	3	GTGGAAAGGACGAAACACCG	GTCTTAGTGAGCCACCGGCG	GTTTTAGAGCTAGAAATAG	**III, IV**
CDH1	4	GTGGAAAGGACGAAACACCG	TCAGAAAGGGCTTTTACACT	GTTTTAGAGCTAGAAATAG	**III, IV**
CDH1	5	GTGGAAAGGACGAAACACCG	GAGACAAGTCGGGGCGGACA	GTTTTAGAGCTAGAAATAG	**III, IV**
REX1	8	GTGGAAAGGACGAAACACCG	TAGCAATACAGTCACATTAA	GTTTTAGAGCTAGAAATAG	**IV**
REX1	10	GTGGAAAGGACGAAACACCG	CCGGGCAGAGAGTGAACGCG	GTTTTAGAGCTAGAAATAG	**IV**
PDX1	1	GTGGAAAGGACGAAACACCG	GCCCCACGTGGTTCAGCCGG	GTTTTAGAGCTAGAAATAG	**III**
PDX1	2	GTGGAAAGGACGAAACACCG	GCCTGGCTGGCCGCACTAAG	GTTTTAGAGCTAGAAATAG	**III**
PDX1	3	GTGGAAAGGACGAAACACCG	AGCAGGTGCTCGCGGGTACC	GTTTTAGAGCTAGAAATAG	**III**
PDX1	4	GTGGAAAGGACGAAACACCG	GTTTGCTGCACACTCCTGAA	GTTTTAGAGCTAGAAATAG	**III**
PDX1	5	GTGGAAAGGACGAAACACCG	GTTTTCGTGAGCGCCCATTT	GTTTTAGAGCTAGAAATAG	**III**
NKX6.1	1	GTGGAAAGGACGAAACACCG	GTAGCGCACTTTGAACAGCT	GTTTTAGAGCTAGAAATAG	**III**
NKX6.1	2	GTGGAAAGGACGAAACACCG	AAACTCTCCGGAGCCAGCCT	GTTTTAGAGCTAGAAATAG	**III**
NKX6.1	3	GTGGAAAGGACGAAACACCG	AGGACGCCTTGTGCAGCCCG	GTTTTAGAGCTAGAAATAG	**III**
NKX6.1	4	GTGGAAAGGACGAAACACCG	CCGAATCTCCACTTTGAAGT	GTTTTAGAGCTAGAAATAG	**III**
NKX6.1	5	GTGGAAAGGACGAAACACCG	GCTCTGCTCTTTCGGTCGCG	GTTTTAGAGCTAGAAATAG	**III**
tetOp	1	GTGGAAAGGACGAAACACCG	GTACCTTCTCTATCACTGAT	GTTTTAGAGCTAGAAATAG	**III**
tetOp	2	GTGGAAAGGACGAAACACCG	GGACTTCTCTATCACTGATA	GTTTTAGAGCTAGAAATAG	**III**
tetOp	3	GTGGAAAGGACGAAACACCG	GGGGAGACGTGCGGCCAGCT	GTTTTAGAGCTAGAAATAG	**III**

8.13. Golden Gate Assembly (III, IV)

Golden Gate assembly was performed as described by Cermak *et al*. [604]. For Golden Gate reaction, gRNA cassettes were amplified with compatible ends using 1 to 5 aggc forward and reverse primers. PCR amplified guide cassettes were concatenated in the Golden Gate reaction into GG-dest backbone (Addgene plasmid #69538). Cloning reaction was performed with 150 ng of GG-dest plasmid and 50 ng of each of the guide PCR fragments. Reaction was run for 50 cycles of 2 min at 37°C, 5 min at 16°C, followed by enzyme inactivation by 20 min at 80°C and transformation into DH5α competent bacteria.

Golden Gate reaction:

T4 ligase buffer (10x) (Thermo Fisher)	2 µl	Thermal cycle:	
T4 DNA ligase (Thermo Fisher)	1 µl	50x	37°C - 2’
Esp3I (Thermo Fisher)	1 µl		16°C - 5’
DTT (10mM) (Promega)	2 µl	1x	80°C - 20’
guide PCR amplicon (up to 5)	x µl (50 ng of each)		
GG-dest plasmid	x µl (150 ng)		
H_2O	x (up to 20 ul)		
Total 20 ul			

Table 4. *Golden Gate primer sequences used in this work*

primer	sequence	sticky end	compatible with	article
1_aggc_Fw	ACTGAATTCGGATCCTCGAG**CGTCTC** A**CCCT**G TAAAACGACGGCCAGT	GGGA	GG-dest plasmid	**III, IV**
1_aggc_Rv	CATGCGGCCGCGTCGACAGATCT**CGTCTC** A**CATG**A GGAAACAGCTATGACCATG	CATG	2_aggc_Fw	**III, IV**
2_aggc_Fw	ACTGAATTCGGATCCTCGAG**CGTCTC** A**CATG**G TAAAACGACGGCCAGT	CATG	1_aggc_Rv	**III, IV**
2_aggc_Rv	CATGCGGCCGCGTCGACAGATCT**CGTCTC** A**GTCC**A GGAAACAGCTATGACCATG	GGAC	3_aggc_Fw	**III, IV**
3_aggc_Fw	ACTGAATTCGGATCCTCGAG**CGTCTC** A**GGAC**G TAAAACGACGGCCAGT	GTCC	2_aggc_Rv	**III, IV**
3_aggc_Rv	CATGCGGCCGCGTCGACAGATCT**CGTCTC** A**CTGG**A GGAAACAGCTATGACCATG	CCAG	4_aggc_Fw	**III, IV**
4_aggc_Fw	ACTGAATTCGGATCCTCGAG**CGTCTC** A**CCAG**G TAAAACGACGGCCAGT	CTGG	3_aggc_Rv	**III, IV**
4_aggc_Rv	CATGCGGCCGCGTCGACAGATCT**CGTCTC** A**AACA**A GGAAACAGCTATGACCATG	TGTT	5_aggc_Fw	**III, IV**
5_aggc_Fw	ACTGAATTCGGATCCTCGAG**CGTCTC** A**TGTT**G TAAAACGACGGCCAGT	AACA	4_aggc_Rv	**III, IV**
5_aggc_Rv	CATGCGGCCGCGTCGACAGATCT**CGTCTC** A**CGTT**A GGAAACAGCTATGACCATG	CGTT	GG-dest plasmid	**III, IV**

8.14. HEK293 Transfection

For transfection, HEK293 cells were seeded on gelatin coated 24-well plates 100 000 cells per well. One day later cells were transfected with 500 ng of DNA using FuGENE HD reagent. For guide testing 500 ng of dCas9 activator plasmid was mixed with 100 to 200 ng of guide cassette PCR amplicon before complexing with the transfection reagent. CRISPR-mediated target gene activation was assessed 3 days after transfection.

9. Results and Discussion

9.1. AAV-mediated Reprogramming (I)

Due to the primarily episomal nature of the rAAV genomes, the system forms a potentially useful tool for inducing transgene-free iPSCs. AAVs are also non-pathogenic and non-toxic, which makes the vector attractive for pluripotent stem cell induction in the light of potential therapeutic applications.

9.1.1. AAV2 Mediated Gene Transduction Induces Colony Formation (I)

To test for AAV mediated gene transfer in pluripotent reprogramming, the reprogramming factors OCT4, SOX2, KLF4 and c-Myc were cloned into AAV vectors under both CMV and CAG promoters. Out of the tested serotypes (AAV2, AAV8 and AAV9) with a GFP control virus, AAV2 virus was found to transduce fibroblasts with the highest efficiency (**I**: **Fig. 1a**). AAV2 was therefore used for virus production of the reprogramming factors and all the reprogramming experiments were done with it. The expression of the transgenes was validated by immunocytochemistry and Western blot (**I**: **Fig. S1**).

Human skin fibroblasts and mouse embryonic fibroblasts were transduced with the reprogramming factor encoding AAVs with varying multiplicities of infection between 20 and 200 000 viral genomes per cell per factor. Colony formation was observed from the transduced fibroblasts, however most of the human fibroblast derived colonies had non-iPSC-like morphologies (**I**: **Fig. S2**). The factors under a CAG promoter induced more robust colony formation than under CMV promoter. This may be due to higher expression of the CAG (or CAGG) than CMV promoter in many cells [605]. Some of the AAV induced human colonies were expanded further and analysed for transgene expression. Particularly OCT4 and c-Myc transgenes were persistently expressed in the clones (**I**: **Fig. S2d**), indicating that persistent expression of only some of the reprogramming factors may have skewed the reprogramming process. This may also be due to problems in efficient transgene delivery, due to insufficient transduction efficiency, resulting in too early loss of some of the transgenic episomes. As the GFP transduction efficiency was at best 70% with the highest MOI tested (**I**: **Fig. 1d**), the fraction of cells containing all of the transgenic reprogramming factors may be too low, especially if multiple transductions are needed. More efficient AAV serotypes or inclusion of the reprogramming factors in fewer multicistronic vectors could help with the factor delivery [350,418].

9.1.2. AAV2 Reprograms Mouse Cells to Pluripotency (I)

The transduction of MEFs with the CAG driven reprogramming factors resulted in the formation of mouse iPSC-like colonies with MOI of 20 000 and 200 000 viral genomes per cell per factor (**I**: **Fig. 1b**). In total, 24 of the resulting miPSC colonies were propagated further and four of the clones were characterised for

pluripotency. These clones stained positive for SSEA-1 and Nanog, showed expression of pluripotent state transcription factors by RT-PCR and formed EBs positive for ectodermal, mesodermal and endodermal markers, demonstrating their pluripotent characteristics (**I**: **Fig. 2**). However, the analysed clones also showed persistent expression of the reprogramming transgenes (**I**: **Fig. 2b**), which may maintain the pluripotent features even if the pluripotent reprogramming may be only partial. Additionally, the persistent transgene expression indicates potential integration of the reprogramming factors.

9.1.3. AAV Reprogrammed Mouse iPSCs Contain Integrated Vector Sequences (I)

The AAV miPSCs were analysed by PCR from genomic DNA and Southern blot to detect potential integration of the reprogramming vectors. This demonstrated integration of multiple copies of the reprogramming vectors in all of the derived miPSC clones (**I**: **Fig. 3**). The fact that miPSC clones contained integrated vector sequences is not completely unexpected due to the high MOI required for colony formation and the estimated integration frequencies of approximately 0.2-1.0 × 10^{-3} per infectious vector genome [428]. Furthermore, the reprogramming process appears to select for clones with the full set of reprogramming factors integrated, possibly due to more stable pluripotent state, as all of the analysed miPSC clones contained integrations of all of the reprogramming factors. Similar results have also been described by Chen *et al*. who reprogrammed mouse adipose derived mesenchymal stem cells with polycistronic tyrosine mutant AAV2 vectors [606]. These miPSCs contained up to 3 integrations per clone of the reprogramming vector. Unlike the CAG driven reprogramming factors, the CMV driven construct used by Chen *et al*. ended up silenced in the resulting miPSC clones. This is likely a result of differences in the behaviour of the promoters, as CMV promoter is known to silence easier in embryonic stem cells than CAG [607]. Due to the preferential integration of the rAAV delivered vectors in the reprogramming process, the system could be utilised to deliver transgenes meant for integration, for example for the generation of secondary reprogrammable inducible mouse cell lines under doxycycline-controlled promoter. This could potentially result in more efficient reactivation of the transgenes than for example lentiviral transgene delivery, as the rAAV backbone is not necessarily targeted to similar extent for silencing by the cells repressive machinery as transgenic retroviral elements [347].

Why do the rAAV vectors integrate with such high frequency in reprogramming? As the genome of rAAV vectors is linear DNA before being circularised by the cell's DNA damage repair machinery, the vectors provide suitable substrates for opportunistic integration into existing genomic double strand breakage sites. The reprogramming process itself has also been linked with increase in DNA damage [244]. Therefore, it is possible that the reprogramming process increases the occurrence of genomic DNA DSB sites, providing more opportunities for the vectors to integrate. The potentially higher incidences of rAAV integration due to DNA DSBs combined with preferential reprogramming of colonies with stable expression of the reprogramming

factors, obviating the need for efficient re-transduction, most likely results in the high occurrence of detected vector integrations in rAAV reprogrammed iPSCs.

9.2. The Effect of Aging on Reprogramming Efficiency (II)

Efficient induction of pluripotency is important for robust generation of iPSC lines. This is particularly important for patient specific disease modelling cell lines where the source cell material may be of variable origin. To investigate the effect of donor age and time of cells in culture on reprogramming efficiency, we reprogrammed skin fibroblasts from 11 donors between 0 and 83 years of age.

9.2.1. Donor Age and Cell Culture Time Correlates Negatively with Reprogramming Efficiency (II)

We first reprogrammed passage 6 fibroblasts of all 11 donors using retroviral delivery of OCT4, SOX2 KLF4 and c-Myc. The reprogramming efficiency was found to decrease by the age of the donor (**II**: **Fig. 1C**). The cells also demonstrated increase in doubling time indicative of slower cell proliferation (**II**: **Fig. 2F,G**). As the retroviral infection is cell proliferation dependent, the efficiency of transgene delivery could be affected by the age dependent decrease in proliferation rate. To this end, the subsequent inductions were made using Sendai viral delivery of the reprogramming transgenes. Next, the combined effect of both donor age and artificial *in vitro* aging by extended culture time were examined. Both increasing age and passage number negatively affected the reprogramming efficiency (**II**: **Fig. 2C,D**). The increase in passage number also correlated with increase in doubling time (**II**: **Fig. 2F**). The high passage samples also demonstrated shortening of telomeres indicative of replicative senescence (**II**: **Fig. 2E**).

It is possible that the age-associated decrease in reprogramming efficiency is not caused by high age *per se* but may be caused by secondary age-associated effects like senescence induced changes in cell proliferation. A recent paper by Lo Sardo *et al*. did not detect effect of donor age on reprogramming efficiency of peripheral blood mononuclear cells (PBMC) by episomal plasmids [608]. The reprogramming of the PBMC population may, thus, not be affected to similar extent by age and cell proliferation as fibroblasts. The problem with our fibroblast experiments and the blood cell reprogramming by Lo Sardo *et al*. is the relatively low number of donor samples used (11 and 16). It is possible that the trend with age, reprogramming efficiency and proliferation could be different if more reprogramming samples were used. This would, however, require high throughput reprogramming of large numbers of cell lines.

9.2.2. Decrease in Reprogramming Efficiency is P21 Dependent (II)

To decipher the mechanism behind the decrease in reprogramming efficiency and cell proliferation, the expression of a set of senescence associated markers was evaluated. P21 expression was found to be significantly upregulated in later passage samples and to correlate with the age of the fibroblasts (**II**: **Fig. 3A-C**). This suggested increased P21 expression as a causative factor in decreasing the reprogramming efficiency. P21 is a well-known factor reducing the reprogramming efficiency. P21 is upregulated as a response to the reprogramming induced cellular stress and downregulation of P21 expression can improve reprogramming efficiency in a cell proliferation dependent manner [214,239]. The effect of P21 on reprogramming efficiency was therefore tested with the aged fibroblasts. Knock down of P21 with siRNA transfection resulted in increase in reprogramming efficiency of late passage fibroblasts and allowed reprogramming of otherwise refractory late passage fibroblasts from old donors (81 and 83 years) (**II**: **Fig. 3G** and **Fig. S3A,B**). Additionally, forced overexpression of P21 in early passage young donor fibroblasts blocked cell proliferation and completely abrogated colony formation (**II**: **Fig. S3E**). This indicated a clear causative role of P21 in the reduction of reprogramming efficiency in the old and late passage fibroblasts.

9.2.3. Cell Proliferation in Late Passage Reprogramming

Cell proliferation rate appears to be a critical defining factor in human fibroblast reprogramming efficiency. The detected upregulation of P21 in late passage and aged fibroblasts is likely linked to the detected shortening of telomeres in the late passage cells, as P21 upregulation commonly results from telomere shortening and dysfunction [609]. Knock out of P21 can bypass the telomere shortening induced replicative senescence in human fibroblasts [610]. It is also possible to immortalise human fibroblasts by overexpression of telomerase. This approach also results in formation of cell populations that can reprogram with constant efficiency over extended period of time in culture [278]. Therefore, when reprogramming aged human fibroblasts, it should be considered whether measures should also be taken to target the telomere-P21 axis, like overexpression of telomerase, ZSCAN4 or cell cycle promoting kinases [216,611]. More in depth understanding of mechanisms in alternative cellular reprogramming approaches that are less dependent on active proliferation, like SCNT or heterokaryons, could also be beneficial for developing more efficient reprogramming methods. For example, the reprogramming of somatic cell nuclei by SCNT can rejuvenate telomeres in Telomerase-deficient mice [612]. As oocytes contain the full set of factors required for reprogramming the epigenome of somatic cell nuclei into totipotent state, SCNT is not as dependent on stochastic cell cycle-dependent events that promote acquisition of pluripotent state in transgenic factor induced pluripotent reprogramming. Therefore, comprehensive targeting of gene regulatory networks, in a manner mimicking the SCNT process, may be beneficial for reprogramming cells in a less stochastic and cell proliferation-dependent manner. Whether cell proliferation is required for induction of pluripotency is

unclear. As the resulting cell population is normally actively dividing, proliferation will need to be promoted in the reprogramming cells. This could maybe be tested by inducing cells into a diapause mimicking state [207,208]. The experiments using heterokaryon formation by cell fusion, which does not required cell division, as a reprogramming tool would suggest that active cell proliferation is not absolutely required for reprogramming to pluripotency if the factor composition is otherwise optimal [613].

As CRISPRa-mediated gene targeting has a high multiplexing capacity, this approach may provide useful also for reducing the cell cycle dependence of late passage fibroblast reprogramming, by promoting comprehensive pluripotency gene activation and reducing the need for stochastic cell cycle-dependent reprogramming events.

9.3. CRISPR-mediated Gene Activation as a Reprogramming Tool (III, IV)

CRISPR-mediated gene activation targets the endogenous gene loci for transcriptional activation. This property is likely beneficial for reprogramming applications, where reliable activation of endogenous gene regulatory networks is required for faithful recapitulation of target cell phenotypes. Particularly, the approach should be useful in transdifferentiation applications as these have been described to commonly have issues with either inefficient activation of the target cell gene expression program or incomplete silencing of the starting cell program [113]. Although CRISPRa is commonly targeted to promoter areas to activate gene transcription, it can also be used to target endogenous gene regulatory elements for epigenetic editing. This allows targeting of particular loci to improve the specificity of resetting the endogenous gene regulatory programs, for example by targeting pluripotency factor binding sites to promote their accessibility. In addition to targeting characterised elements, the tool can be used to study the function of uncharacterised gene regulatory elements in reprogramming. Therefore, CRISPRa is a highly versatile tool and suitable for various reprogramming applications. Additional benefit of using CRISPRa-mediated gene activation for reprogramming is the high multiplexing capacity of the system. As the genes are targeted with short RNA molecules, the system can be used to maximise the number of targeted reprogramming factors, making it easier to promote cooperativity of these factors in the reprogramming process.

9.3.1. CRISPRa Promotes Activation of Pluripotent Reprogramming Factors (III, IV)

The development of a CRISPRa reprogramming system was initiated by building catalytically deactivated Cas9 activator constructs by fusion of the dCas9 with various numbers of VP16 acidic activator peptide repeats (**III**: **Fig. 1A**). These were tested in the activation of endogenous *OCT4* locus by targeting the dCas9 activator constructs to the *OCT4* promoter proximal sequences. An increase in the transcriptional output of *OCT4* correlated with the increasing numbers of the VP16 repeats (**III**: **Fig. 1B,C**). This approach was also applied to

other pluripotent reprogramming factors, which could all be activated to various extent (**III**: **Fig. 1E-G**). The pluripotency factor promoter guide composition was further optimised by testing each guide separately and incorporating the best working guides into one vector for easier delivery into cells (**IV**: **Fig. 2b,c**). These guides were tested in HEK293 with dCas9VPH activator, containing additional P65-HSF1 activator domains in addition to the VP16 peptides, and found to mediate efficient activation of the canonical reprogramming factors (**IV: Fig. 2d**). In HFFs these guides mediated efficient activation of *OCT4* and *SOX2,* whereas *KLF4*, *MYC* and *LIN28A* demonstrated poor activation (**IV**: **Fig. 2d**). The targeting of the gene promoters with pools of five guides was more efficient in activating the genes than any single guides (**IV**: **Fig. 2b**), similarly to what has been previously described [510–513]. This demonstrated that CRISPRa can be used to efficiently target human pluripotency factors for transcriptional activation and that the reprogramming factor targeting guides can be combined for multiplexed targeting of the reprogramming factors genes with simpler vector constructs.

9.3.2. CRISPRa Replaces Transgenic OCT4 in Reprogramming (III, IV)

To test if the gene activation level was sufficient to have a practical impact on reprogramming, we replaced the transgenic OCT4 encoding plasmid in the episomal reprogramming of human fibroblasts with a dCas9VP192 (12 × VP16 activator peptide repeat) encoding plasmid and another plasmid containing five *OCT4* promoter targeting guides. This resulted in efficient activation of the endogenous *OCT4* transcription, which was sufficient for successful reprogramming of both neonatal and adult human skin fibroblasts with transgenic SOX2, KLF4, L-MYC, LIN28A and CRISPRa-mediated *OCT4* activation (**III**: **Fig. 2D-H** and **Fig. S2**). Reprogramming of iPSC-derived neural stem cells (NSC) to pluripotency was also possible with dCas9 activator-mediated targeting of *OCT4* promoter (**IV**: **Fig. 1**), similar to what has been described for transgenic OCT4 overexpression [163]. This demonstrated that full CRISPRa reprogramming of human cells was possible by *OCT4* activation from specific cell populations. NSCs are developmentally close to pluripotent state and can be reprogrammed easier than more distant somatic cell types, the reprogramming of skin fibroblasts would thus be expected to require activation of additional reprogramming factors.

9.3.3. Guide Optimization Promotes Full CRISPRa Reprogramming (IV)

Our initial attempts at reprogramming human skin fibroblasts with CRISPRa were unsuccessful with combination of vectors targeting *OCT4*, *SOX2*, *KLF4*, *LIN28A* and *NANOG* for activation with plasmids encoding five guides per gene promoter. As the problem may have been in any of the components of the gene activation system, i.e. inefficient dCas9 activator or guide composition, we tried to improve all the possible aspects of the system. Particularly, novel dCas9 activator constructs were built, containing additional P65-HSF1 and P300 core domains [518,550], to try and improve the gene activation efficiency. Reprogramming factor targeting guide composition was also optimized to include only one to three of the best performing

guides per gene (**IV**: **Fig. 2c**). The activator was targeted to *OCT4*, *SOX2*, *KLF4*, *LIN28A* and *MYC* promoters, and an EGA enriched Alu-motif (EEA-motif) (**IV**: **Fig. 1d**), commonly found near the early human embryo genome activation expressed gene promoters [595]. The combination of all these modifications resulted in the formation of the first CRISPRa-induced iPSC colonies.

Once the CRISPRa reprogramming was initially established as a method for human fibroblast reprogramming, the effect of each of the separate components on the reprogramming efficiency was further analysed.

The reprogramming efficiency was found to be the highest with the dCas9VP192 activator (**IV**: **Fig. 3g**). This was unexpected, as the reprogramming efficiency should generally increase upon higher reprogramming factor expression levels, and the more complex activator domains would be expected to promote more efficient target gene activation. Particularly, the VP192-P300 core fusion activation domain showed more efficient activation of *OCT4* (**IV**: **Fig. S4c**). It is possible that the enzymatic modifications mediated by the P300 core domain may have some detrimental global effect on the cells, as suggested by increase in histone 3 tail acetylation by the VP192-P300 construct (**IV**: **Fig. S4e,f**). Alternatively, detrimental local epigenetic effects in the promoter proximal areas of the genes targeted may be affecting the gene activation. To decipher what is happening at the targeted areas, the starting cell epigenetic status would need to be correlated with the epigenetic status of the areas at the target cell type, and the effect of the activation domains to the targeted areas would need to be analysed. Additionally, the endogenous loci may be targeted by the endogenous somatic cell factors, which may interfere with the artificial activation of the sites, for example by repressing the genes intended for activation. Initiating the reprogramming process can further cause stress to the cells, which normally results in reduction of cell proliferation by P53 and P21 mediated pathway. This may also result in downregulation of the endogenous MYC levels by P53 [614,615]. The interaction of the targeted genes, the somatic cell transcriptional machinery and the epigenetic state of the targeted genes may thus need to be taken into consideration when designing optimal CRISPRa-based reprogramming schemes.

The guide composition included in the first successful CRISPRa pluripotent reprogramming scheme included only one guide each for MYC and KLF4. The activation of these two genes was inefficient in HFFs, possibly due to already high expression in the starting cells (**IV**: **Fig. 2d**). Therefore, the addition of an extra plasmid with KLF4 and MYC guides, and a second reprogramming factor containing plasmid with two LIN28A guides changed to KLF4 and MYC guides were tested for reprogramming. Both of these approaches worked better than the initial reprogramming guide plasmid (**IV**: **Fig. S5**), indicating that efficient *KLF4* and *MYC* activation may be a limiting step in the CRISPRa-mediated reprogramming. As the number of possible guide combinations for gene activation in reprogramming is very large, it may be more practical for the near future to select the best functioning CRISPRa reprogramming guides functionally by library screening rather than try to rationally design the guides with our current limited understanding of the CRISPR-mediated gene activation system.

The inclusion of EEA-motif targeting guides was crucial for the efficient reprogramming with the CRISPRa method. The inclusion of these guides in the reprogramming cocktail increased the CRISPRa reprogramming efficiency of neonatal human skin fibroblasts by ten to thirty fold (**IV**: **Fig. 3g**), and was required for the reprogramming of the tested adult skin fibroblasts. Our initial hypothesis was that since this motif was enriched near the early embryo expressed genes, it would function upstream of the acquisition of pluripotent phenotype in the embryonic cells. However, it is not yet clear whether this is a mechanism that contributes to the effect of the EEA-motif targeting in the CRISPRa-mediated pluripotent reprogramming.

9.3.4. Conserved Alu-motif Targeting Promotes Pluripotency Factor Activation (IV)

To gain insight into the function of the EEA-motif targeting in CRISPRa reprogramming, we did RNA sequencing of the CRISPRa reprogramming cell populations with and without the EEA-motif targeting guides. This revealed EEA-motif targeting dependent set of genes at the early half of reprogramming (**IV**: **Fig. 4d**), and pluripotency associated set of genes at the latest time point (**IV**: **Fig. 4c**), which was dependent on both reprogramming factor targeting guides and EEA-motif targeting guides. The set of genes higher expressed at the early time point samples in an EEA-motif guide dependent manner also had EEA-guide 1 target sites enriched in their proximal sequences, indicative of a direct activation effect in an EEA-guide target dependent manner. This set did not contain any clear pluripotency associated factors and may therefore be more dependent on the combined function of EEA-motif targeting guides and somatic factors. The later time point samples did not show similar enrichment for EEA-motif targeting guides. Therefore, the later time point target genes are likely to be more prominently affected by the reprogramming process and possibly aided by the EEA-motif targeting. Alu and B1 sequences have, however, been reported to be enriched near pluripotency factor promoters [596], which may contribute to the EEA-motif targeting effect. It is possible that this effect is not detected in the RNA sample pools due to fibroblast background and possible transient effect of the EEA-motif targeting.

To determine if the late stage upregulated reprogramming factor guide and EEA-motif guide dependent genes had a role in the EEA-motif targeting effect on CRISPRa reprogramming, a set of day 12 upregulated genes were tested for their impact on reprogramming efficiency. Of the tested genes, transgenic expression of NANOG and REX1 increased the CRISPRa reprogramming efficiency (**IV**: **Fig. 5b**). Both of these genes were also more efficiently activated in HEK293 by dCas9VP192 targeted to their promoter proximal areas in the presence of EEA-motif targeting guides (**IV**: **Fig. 5c**). This indicated that *NANOG* and *REX1* are two likely targets that are mediating the EEA-motif targeting effect on pluripotent reprogramming, as their activation is improved by EEA-motif guides and their expression improves reprogramming efficiency. However, *NANOG* and *REX1* are unlikely to be the only target genes of the EEA-motif guides due to the high numbers of these

sites in the human genome. Identification of additional EEA-motif targeting affected genes will likely require either a more efficient reprogramming system or purification of reprogramming intermediate populations.

9.3.5. Inducible CRISPRa (III, IV)

To gain more control over the timing of the CRISPRa-mediated gene activation, we constructed dCas9 activator fused with DHFR-derived trimethoprim (TMP) stabilised degradation domain [616] (**III**: **Fig. 3**). Stable expression of this construct gave TMP dependent activation of targeted genes in HEK293 and human skin fibroblast cells (**III**: **Fig. 3**). The construct did, however, demonstrate some level of leakiness, possibly due to incomplete degradation of the fused dCas9 activator protein and the strong VP192 activator domain. This leakiness could be controlled by putting the destabilised dCas9 activator under a doxycycline (DOX) inducible promoter. The combination of both TMP and DOX inducible systems were sufficient to reduce the leakiness compared to both systems separately (**III**: **Fig. 3b,c**). The use of these systems separately allowed separate temporal control of the expression of two different sets of factors (**III**: **Fig. 4f-h**).

To reprogram human skin cells with the inducible CRISPRa system, the TetON-DDdCas9VP192 activator was combined with the reprogramming factor targeting guides and the EEA-motif targeting guides in PiggyBac transposon vector system. This way neonatal human foreskin fibroblast cells could be selected after activator and guide delivery, and reprogramming could be induced by addition of TMP and DOX in the culture medium (**IV**: **Fig. S8**). This system could also be used to produce stable pluripotent stem cell clones, which could be re-reprogrammed after differentiation (**IV**: **Fig. S8g-i**). This inducible CRISPRa reprogramming system will likely be a useful tool for further investigation into the mechanisms behind the CRISPRa-mediated induction of pluripotency.

9.3.6. What is the Mode of Action of EEA-motif Targeting?

The EEA-motif is located in a conserved Alu consensus sequence in the left arm of the Alu element. The guide target sites targeted in this motif are located in between the A and B-boxes of the Pol III promoter of the Alu element and contain multiple characterised binding sites for a number of different transcription factors nearby [593]. The elements themselves, particularly a specific subset of B1-elements, have been implicated to have insulator-like function that is activated by active transcription of the motif promoter [588]. The chromatin looping associated factor CTCF is recruited to the activated B1 elements in a PARP-1 dependent manner leading to the establishment of insulation at these loci. In addition to CTCF, the Alu consensus motif sequences are known to contain binding sites for other chromatin architecture associated factors like YY1 [79,594], and TFIIIC, a transcription factor important for Pol III RNA polymerase initiation complex. TFIIIC and its interaction with cohesin have been linked to chromatin insulation at tRNA gene B-box sequences [617], and

TFIIIC and condensin II have been shown to support expression of TAD boundary associated genes [618]. Therefore, it is quite likely that the function of the EEA-motif and its targeting in reprogramming may be associated with the control of chromatin 3D structure and the role of DNA conformation in gene transcriptional control. To gain a better idea of the function of the motif, chromatin conformation capture and proteomics analysis would need to be performed. This would also help elucidate the molecular mechanisms governing the EEA-motif function. This may, however, be additionally complicated by the context dependence of the possible chromatin interaction function, as the transcriptional effect of targeting these sites may be dependent on the presence of additional factors that cause transcriptional changes to the associated genes, i.e. pluripotent reprogramming factors. A role of Alu-motifs and TFIIIC in controlling chromatin 3D architecture is implied by the contribution of these elements to the activity dependent relocation of neuronal genes to transcription factories [590]. Therefore, EEA-motif targeting may be promoting similar function in pluripotency-associated genes upon their activation by the reprogramming factors. It is also not clear whether the EEA-motif targeting effect by dCas9 is mediated by opening of the sequences near these elements or blocking of the binding sites in the guide target sites. Theoretically, both of these options could be possible. Blocking of, for example, TFIIIC could promote more efficient relocation of these sites to transcription factories (see **Figure 12** model), or blocking of insulator function could promote enhancer-promoter interaction over the insulated regions. On the other hand, opening of these regions could provide more accessibility to, for example, YY1 to mediate intra TAD interactions to promote active transcription. Activation of these regions could also promote activation of the insulator function for establishing pluripotency associated insulated boundaries. Our results by targeting the EEA-motif with dCas9 in HEK293, demonstrated a slight increase in ATAC-seq accessible regions at EEA-guide 1 sites (**IV: Fig. 6f**). This suggest a chromatin opening effect of the dCas9 protein. However, it does not rule out the possibility of blocking function.

9.3.7. Model for Targeted Gene Activation in Reprogramming

Cellular reprogramming by targeted activation of endogenous genes requires consideration of additional issues that may not be equally relevant for transgenic factor mediated reprogramming. In transgenic reprogramming, high expression of the reprogramming factors is usually sufficient to initiate the reprogramming process. As the CRISPR reprogramming is driven by the transgenic dCas9 activator complex, the function of the dCas9 and how it mediates the activation of its target genes are crucial for successful reprogramming. The reprogramming process normally requires simultaneous activation of a set of transcription factors. Therefore, the dCas9 activator has to function robustly at different target genes. It is not currently well known how the dCas9 activator targeting will affect the targeted genomic loci, particularly in a process where the transcriptional mechanisms that are maintaining the target gene expression may

depend on the changing cellular phenotype. In the case of reprogramming, the complexes that are promoting efficient transcription in somatic cells may be different from the ones in pluripotent cells, and optimisation of dCas9 activator function in somatic cells may not produce best result for inducing robust expression in pluripotent state. To understand this better, the mechanisms that are controlling the transcription will need to be studied in more detail. Considering the transcriptional activation of a gene from the phase separation model perspective [65], the possible differences in the factors controlling this phenomenon in somatic cells and in pluripotent cells may contribute to the stable transcription of the targeted genes. This may be part of the issue with the observed decrease in reprogramming efficiency with the P65-HSF1 containing activator domains (**IV**), but this issue would need to be further investigated. Additionally, targeting of other areas, such as enhancers and super enhancers, would be expected to further stabilise the expression of the artificially activated genes, as this could contribute to more efficient phase separation of the targeted loci.

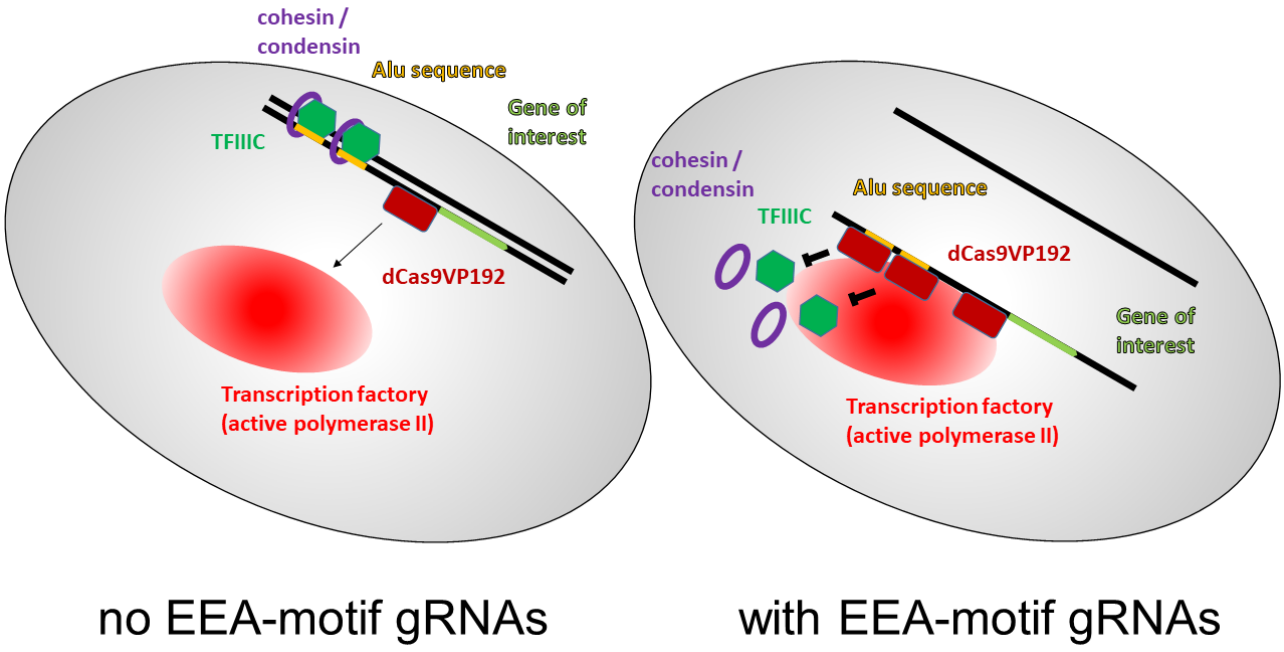

Figure 12. *A schematic representation of a model for EEA-motif function in CRISPRa-mediated gene activation. Increased numbers of EEA-motifs near pluripotency genes may promote more efficient translocation and phase separation of these loci into transcription factories in the presence of EEA-motif targeting gRNAs. Binding of dCas9 into the Pol III promoter of the Alu left arm may interfere with TFIIIC function and chromatin cohesion, enabling changes in chromatin 3D architecture that promote active transcription.*

Conventionally, the transcriptional activation of the targeted genes has been considered to occur via the recruitment of the mediator and other Pol II recruiting complexes to the targeted promoters by the dCas9

fused transactivation domain. In this case, the activation efficiency would be expected to rise with the addition of extra activation domains. This is not, however, what has been observed. Instead the activation efficiency seems to plateau after a certain point [522]. It is possible that the recruitment of the dCas9 activator to multiple sites near a gene will promote more efficient activation of the gene than extra activation domains. This appears to also be supported by the increased activation efficiency of targeted genes by pooled guide RNAs (**IV**: **Fig. 2b**). Accordingly, targeted CRISPR-mediated gene activation may therefore work by translocation of the physical gene locus into the actively transcribed nuclear areas, i.e. promoting the translocation of the genes to transcription factories. This would be mechanistically analogous to what has been described for activity dependent relocation of neuronal genes to transcription factories. Thus, it is also possible that the EEA-motif targeting in reprogramming mimics the function of the SINE motifs as described by Crepaldi *et al*. [590]. Mechanistically this could mean that EEA-motif targeting may interfere with the TFIIIC / condensin mediated DNA packaging and promote more efficient translocation of the pluripotency factor loci into transcription factories due to high number of dCas9 targeted EEA-motifs near these genes and simultaneous targeting of these genes for activation by the reprogramming factors. The exact mechanisms of the CRISPR reprogramming process will still need to be further studied. This will optimally require high efficiency inducible reprogramming systems, where reprogramming can be initiated in a large percentage of the cell population, in order to get sufficient numbers of reprogramming cells needed for comprehensive analysis of the CRISPRa reprogramming process. A model for EEA-motif mediated promotion of CRISPRa gene activation in reprogramming is presented in **Figure 12**.

10. Conclusions and Future Perspectives

Induction of pluripotency has been a revolutionising technology for biomedical research. The extensive amount of research invested in this technology during the past decade has led to rapid development of the cellular reprogramming field. The advent of the CRISPR era is likely to revolutionise the field further. With the capability to produce sequence specific genetic and epigenetic changes in chromatin and DNA, our ability to produce high quality reprogrammed and gene edited cells for research and potentially therapeutic applications will improve greatly.

In this thesis we have demonstrated that rAAV vectors can be used to reprogram mouse embryonic fibroblasts to pluripotency. However, this method is associated with high frequency of vector integration, which limits its use for derivation of integration free iPSCs. This work highlights the potential of rAAV vector integration in some applications and urges consideration of the vector integration risk. If vector integration is not an issue for the planned application, the method can potentially be used for example for generation of iPSCs for secondary reprogramming experiments. As the integration of rAAV vectors is opportunistic, the vector integration could also be used for detecting DNA DSB sites generated during the reprogramming process.

Comparison of reprogramming of aged and long term cultured human skin fibroblasts demonstrated the role of P21 in limiting the efficiency of reprogramming of old donor fibroblasts and cells that have been extensively expanded in culture. In order to improve the reliability of reprogramming from variable sources of human fibroblasts, P21 inhibition should be taken into consideration. Particularly, the inhibition of P21 instead of P53 could be useful, as the P21 levels were already upregulated in the aged fibroblasts. This work was performed with relatively low number of samples and therefore further work with higher sample sizes should inform us more of the effect of aging on reprogramming efficiency. Moreover, inclusion of alternative cell types should give a better idea of the general applicability of P21 inhibition. Other effects than cell proliferation should also be taken into consideration. It is likely that P21 upregulation and inhibition of cell proliferation are not the only effects of aging on reprogramming and that additional factors, such as epigenetic changes, may contribute to the process. In order to avoid the negative effect of decreased proliferation more advanced reprogramming methods that are less dependent on cell proliferation need to be developed.

By optimising dCas9 transcriptional activators, guide composition for endogenous reprogramming factor targeting and targeting of conserved Alu sequences, we were able to demonstrate CRISPRa-mediated reprogramming of human somatic cells to pluripotency. This is the first time functional reprogramming of human cells to pluripotency using CRISPRa has been described. CRISPRa-mediated reprogramming

technology is likely to have a great impact on the cellular reprogramming field. The system is unique as it allows targeting of endogenous loci for activation and simultaneous targeting of a large sets of genomic areas. This is not only useful for induction of pluripotency, but potentially also for transdifferentiation. It is likely that the CRISPRa method is not going to be used only by itself, but to be combined with transgenic transcription factors, RNAi and small molecular compounds to best guide the reprogramming process. This will aid in comprehensive targeting of the transcriptional programs for maximising the accuracy of resetting the cellular phenotype. One problem with the current approaches for CRISPRa reprogramming is the complexity of the guide targeting. Many possible target genes and guide sites in promoter and enhancer regions make it difficult to predict optimal reprogramming combinations. Therefore, for future development of CRISPRa reprogramming, it may be best to apply guide screens to select best working guides in a functionally relevant setting. Our CRISPRa reprogramming method relies on the targeting of conserved Alu-motifs for efficient reprogramming. The function of these elements and their role in reprogramming and development has not yet been thoroughly studied. Therefore, deciphering the molecular function of these motifs is important for understanding their role in reprogramming and development. The targeting of these loci may also turn out to be useful for other reprogramming applications, such as SCNT or CRISPRa-mediated transdifferentiation. As transposable elements have restructured the gene regulatory networks during evolution, and commonly demonstrate tightly temporally controlled expression patterns, the possible role of other transposable elements in cellular reprogramming should be further studied. The CRISPRa reprogramming methods described in this thesis offer a great tool for that.

In conclusion, this thesis work provides a number of novel tools and insights for improving the pluripotent reprogramming process of human cells. This will likely result in better quality of reprogrammed cells, improving the cellular models used for research purposes and potentially providing improved reprogrammed cell populations for future therapeutic applications with closer resemblance to their *in vivo* counterparts.

11. Acknowledgements

This PhD thesis work was carried out between 2011 and 2018 at the Biomedicum Stem Cell Center, Molecular Neurology Research Program, University of Helsinki. I would like to thank the director of the Research Programs Unit, Henna Tyynismaa, as well as the director of the Molecular Neurology Research Program, Anu Wartiovaara, for providing excellent research facilities.

This work has been supported by the Doctoral Programme in Biomedicine, the ESTOOLS Consortium under the Sixth Research Framework Program of the European Union, Academy of Finland, Sigrid Jusélius Foundation, 3i Regeneration project (TEKES Large Strategic Research Opening), and Jane and Aatos Erkko Foundation.

I am most grateful to my supervisor professor Timo Otonkoski for providing me with the opportunity to work with the fascinating topic of pluripotent reprogramming. Your continuous support has enabled me to grow as a researcher and helped foster the development independent research ideas. I have all the time felt that you have believed in me even though the thesis progress has been slow at times. I would also like to thank my co-supervisor Ras Trokovic. Your good sense of humour has made it more fun to work at the lab. This has been important in creating a good working environment.

My thesis committee members professor Olli Silvennoinen and assistant professor Samer Hussein are thanked for their support and encouragement during my thesis project.

Associate professor Noora Kotaja and docent Einari Niskanen are thanked for the review of this thesis and their constructive comments. I would also like to thank doctor Jacob Hanna for accepting invitation to act as my opponent.

I would like to thank all the collaborators and their lab members who have contributed to this work. Particularly, I would like to than Kari and Andrei for the aid in AAV production, Taneli and Ken for the aid in analysing the aging fibroblast phenotype, Kirmo for aid with the CRISPR/Cas9 work and Juha, Shintaro and Kaarel for the Alu targeting and CRISPR reprogramming work.

I am really grateful to all my current and former lab mates at the Biomedicum Stem Cell Center. It has been a pleasure to work with you all these years. Particularly, I want to thank Diego for all the help with the CRISPR stuff. The CRISPR reprogramming would never have become a thing without all the after work brainstorming and science discussions with a side of some productive procrastination. I would also like to thank Sande for sharing the enthusiasm about all the obscure embryo biology stuff, and Yuval for sharing the hype about all the CRISPR stuff. All my fellow grad students, post docs and undergrads: Milla, Elina, Sanna, Väinö, Hazem, Pauliina, Eeva-Mari, Maxim, Jonna, Vikash, Tom, Laura, Anna, Sami, Rocio, Minttu, Masha, Elina, Aki,

Karoliina, Kaisa, Suvi and others. Thank you for all the help with everything, and of course all the parties and cakes. You guys create an amazing lab spirit that makes it fun to come to work. I would also like to thank all the staff scientists and technical personnel in the lab. Heli, Solja, Anni, Eila, Jarkko, Jaan and others, you guys keep the lab running and without you not much would actually work.

I would like to acknowledge all my friends and relatives who have actively tried to drag me out of the grim darkness of the laboratory into the eye burning sunlight. Your effort is truly appreciated. And finally, I would like to thank my siblings, Noora and Jori, my mom and dad, and my grandparents for being there for me throughout my studies. You have always been understanding and supportive of my “mitäsenyttaasolikaan mitä sä opiskelit” studies. I believe academic success stems from the home and I have always felt that you have appreciated and supported my efforts. Thank you for that.

12. References

1. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126,** 663–676 (2006).
2. Mandai, M. *et al.* Autologous Induced Stem-Cell–Derived Retinal Cells for Macular Degeneration. *N. Engl. J. Med.* **376,** 1038–1046 (2017).
3. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci.* **109,** E2579–E2586 (2012).
4. Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (80-.).* **337,** 816–821 (2012).
5. Thomson, J. a. Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science (80-.).* **282,** 1145–1147 (1998).
6. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292,** 154–156 (1981).
7. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci.* **78,** 7634–7638 (1981).
8. Kanatsu-Shinohara, M. *et al.* Generation of Pluripotent Stem Cells from Neonatal Mouse Testis. *Cell* **119,** 1001–1012 (2004).
9. Golestaneh, N. *et al.* Pluripotent Stem Cells Derived From Adult Human Testes. *Stem Cells Dev.* **18,** 1115–1125 (2009).
10. Tachibana, M. *et al.* Human Embryonic Stem Cells Derived by Somatic Cell Nuclear Transfer. *Cell* **153,** 1228–1238 (2013).
11. Noggle, S. *et al.* Human oocytes reprogram somatic cells to a pluripotent state. *Nature* **478,** 70–75 (2011).
12. Nichols, J. & Smith, A. Naive and Primed Pluripotent States. *Cell Stem Cell* **4,** 487–492 (2009).
13. Wu, J. *et al.* An alternative pluripotent state confers interspecies chimaeric competency. *Nature* **521,** 316–321 (2015).
14. Macfarlan, T. S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487,** 57–63 (2012).
15. Yang, Y. *et al.* Derivation of Pluripotent Stem Cells with In Vivo Embryonic and Extraembryonic Potency. *Cell* **169,** 243–257.e25 (2017).
16. Yang, J. *et al.* Establishment of mouse expanded potential stem cells. *Nature* **550,** 393–397 (2017).
17. Choi, Y. J. *et al.* Deficiency of microRNA miR-34a expands cell fate potential in pluripotent stem cells. *Science (80-.).* **355,** eaag1927 (2017).
18. Tonge, P. D. *et al.* Divergent reprogramming routes lead to alternative stem-cell states. *Nature* **516,** 192–197 (2014).
19. van Oosten, A. L., Costa, Y., Smith, A. & Silva, J. C. R. JAK/STAT3 signalling is sufficient and dominant over antagonistic cues for the establishment of naive pluripotency. *Nat. Commun.* **3,** 817 (2012).
20. Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519–523 (2008).
21. Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic

stem cells. *Nature* **448,** 196–199 (2007).

22. Brons, I. G. M. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448,** 191–195 (2007).

23. Vallier, L., Alexander, M. & Pedersen, R. A. Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J. Cell Sci.* **118,** 4495–509 (2005).

24. Ludwig, T. E. *et al.* Derivation of human embryonic stem cells in defined conditions. *Nat. Biotechnol.* **24,** 185–187 (2006).

25. Chen, G. *et al.* Chemically defined conditions for human iPSC derivation and culture. *Nat. Methods* **8,** 424–429 (2011).

26. Singh, A. M. *et al.* Signaling Network Crosstalk in Human Pluripotent Cells: A Smad2/3-Regulated Switch that Controls the Balance between Self-Renewal and Differentiation. *Cell Stem Cell* **10,** 312–326 (2012).

27. Wang, L. *et al.* Self-renewal of human embryonic stem cells requires insulin-like growth factor-1 receptor and ERBB2 receptor signaling. *Blood* **110,** 4111–4119 (2007).

28. Ho, L. *et al.* ELABELA Is an Endogenous Growth Factor that Sustains hESC Self-Renewal via the PI3K/AKT Pathway. *Cell Stem Cell* **17,** 435–447 (2015).

29. Göke, J., Chan, Y.-S., Yan, J., Vingron, M. & Ng, H.-H. Genome-wide Kinase-Chromatin Interactions Reveal the Regulatory Network of ERK Signaling in Human Embryonic Stem Cells. *Mol. Cell* **50,** 844–855 (2013).

30. Tee, W.-W., Shen, S. S., Oksuz, O., Narendra, V. & Reinberg, D. Erk1/2 Activity Promotes Chromatin Features and RNAPII Phosphorylation at Developmental Promoters in Mouse ESCs. *Cell* **156,** 678–690 (2014).

31. Robinson, C. & Lowe, M. Mechanisms and Developmental Roles of Promoter-proximal Pausing of RNA Polymerase II. *J. Stem Cell Res. Ther.* **06,** (2016).

32. Williams, L. H. *et al.* Pausing of RNA Polymerase II Regulates Mammalian Developmental Potential through Control of Signaling Networks. *Mol. Cell* **58,** 311–322 (2015).

33. Estarás, C., Benner, C. & Jones, K. A. SMADs and YAP Compete to Control Elongation of β-Catenin:LEF-1-Recruited RNAPII during hESC Differentiation. *Mol. Cell* **58,** 780–793 (2015).

34. Beyer, T. A. *et al.* Switch Enhancers Interpret TGF-β and Hippo Signaling to Control Cell Fate in Human Embryonic Stem Cells. *Cell Rep.* **5,** 1611–1624 (2013).

35. Bao, S. *et al.* Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. *Nature* **461,** 1292–1295 (2009).

36. Guo, G. *et al.* Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development* **136,** 1063–1069 (2009).

37. Ye, S., Li, P., Tong, C. & Ying, Q.-L. Embryonic stem cell self-renewal pathways converge on the transcription factor Tfcp2l1. *EMBO J.* **32,** 2548–2560 (2013).

38. Silva, J. *et al.* Nanog Is the Gateway to the Pluripotent Ground State. *Cell* **138,** 722–737 (2009).

39. Festuccia, N. *et al.* Esrrb Is a Direct Nanog Target Gene that Can Substitute for Nanog Function in Pluripotent Cells. *Cell Stem Cell* **11,** 477–490 (2012).

40. Hanna, J. *et al.* Metastable Pluripotent States in NOD-Mouse-Derived ESCs. *Cell Stem Cell* **4,** 513–524 (2009).

41. Yang, J. *et al.* Stat3 Activation Is Limiting for Reprogramming to Ground State Pluripotency. *Cell Stem*

Cell **7,** 319–328 (2010).

42. Guo, G. & Smith, A. A genome-wide screen in EpiSCs identifies Nr5a nuclear receptors as potent inducers of ground state pluripotency. *Development* **137,** 3185–3192 (2010).

43. Gillich, A. *et al.* Epiblast Stem Cell-Based System Reveals Reprogramming Synergy of Germline Factors. *Cell Stem Cell* **10,** 425–439 (2012).

44. Ware, C. B. *et al.* Derivation of naive human embryonic stem cells. *Proc. Natl. Acad. Sci.* **111,** 4484–4489 (2014).

45. Buecker, C. *et al.* A Murine ESC-like State Facilitates Transgenesis and Homologous Recombination in Human Pluripotent Stem Cells. *Cell Stem Cell* **6,** 535–546 (2010).

46. Li, W. *et al.* Generation of Rat and Human Induced Pluripotent Stem Cells by Combining Genetic Reprogramming and Chemical Inhibitors. *Cell Stem Cell* **4,** 370 (2009).

47. Hanna, J. *et al.* Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. *Proc. Natl. Acad. Sci.* **107,** 9222–9227 (2010).

48. Wang, W. *et al.* Rapid and efficient reprogramming of somatic cells to induced pluripotent stem cells by retinoic acid receptor gamma and liver receptor homolog 1. *Proc. Natl. Acad. Sci.* **108,** 18283–18288 (2011).

49. Takashima, Y. *et al.* Resetting Transcription Factor Control Circuitry toward Ground-State Pluripotency in Human. *Cell* **158,** 1254–1269 (2014).

50. Qin, H. *et al.* YAP Induces Human Naive Pluripotency. *Cell Rep.* **14,** 2301–2312 (2016).

51. Zimmerlin, L. *et al.* Tankyrase inhibition promotes a stable human naïve pluripotent state with improved functionality. *Development* dev.138982 (2016). doi:10.1242/dev.138982

52. Xu, Y. *et al.* Revealing a core signaling regulatory mechanism for pluripotent stem cell survival and self-renewal by small molecules. *Proc. Natl. Acad. Sci.* **107,** 8129–8134 (2010).

53. Chan, Y.-S. *et al.* Induction of a Human Pluripotent State with Distinct Regulatory Circuitry that Resembles Preimplantation Epiblast. *Cell Stem Cell* **13,** 663–675 (2013).

54. Valamehr, B. *et al.* Platform for Induction and Maintenance of Transgene-free hiPSCs Resembling Ground State Pluripotent Stem Cells. *Stem Cell Reports* **2,** 366–381 (2014).

55. Gafni, O. *et al.* Derivation of novel human ground state naive pluripotent stem cells. *Nature* (2013). doi:10.1038/nature12745

56. Theunissen, T. W. W. *et al.* Systematic Identification of Defined Conditions for Induction and Maintenance of Naive Human Pluripotency. *Cell Stem Cell* **15,** 471–487 (2014).

57. Guo, G. *et al.* Naive Pluripotent Stem Cells Derived Directly from Isolated Cells of the Human Inner Cell Mass. *Stem Cell Reports* **6,** 437–446 (2016).

58. Liu, X. *et al.* Comprehensive characterization of distinct states of human naive pluripotency generated by reprogramming. *Nat. Methods* **14,** 1055–1062 (2017).

59. Choi, J. *et al.* Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. *Nature* **1,** 1–7 (2017).

60. Yagi, M. *et al.* Derivation of ground-state female ES cells maintaining gamete-derived DNA methylation. *Nature* 20–22 (2017). doi:10.1038/nature23286

61. Voss, S. D., Schlokat, U. & Gruss, P. The role of enhancers in the regulation of cell-type-specific transcriptional control. *Trends Biochem. Sci.* **11,** 287–289 (1986).

62. Allen, B. L. & Taatjes, D. J. The Mediator complex: a central integrator of transcription. *Nat. Rev.*

Mol. Cell Biol. **16,** 155–166 (2015).

63. Hnisz, D. *et al.* Super-Enhancers in the Control of Cell Identity and Disease. *Cell* **155,** 934–947 (2013).

64. Whyte, W. a *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153,** 307–19 (2013).

65. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A Phase Separation Model for Transcriptional Control. *Cell* **169,** 13–23 (2017).

66. Jackson, D. A., Hassan, A. B., Errington, R. J. & Cook, P. R. Visualization of focal sites of transcription within human nuclei. *EMBO J.* **12,** 1059–65 (1993).

67. Iborra, F. J., Pombo, A., Jackson, D. a & Cook, P. R. Active RNA polymerases are localized within discrete transcription 'factories' in human nuclei. *J. Cell Sci.* **109 (Pt 6),** 1427–36 (1996).

68. Sutherland, H. & Bickmore, W. A. Transcription factories: gene expression in unions? *Nat. Rev. Genet.* **10,** 457–466 (2009).

69. Strom, A. R. *et al.* Phase separation drives heterochromatin domain formation. *Nature* **547,** 241–245 (2017).

70. Hnisz, D., Day, D. S. & Young, R. A. Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. *Cell* **167,** 1188–1200 (2016).

71. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159,** 1665–1680 (2014).

72. Nichols, M. H. & Corces, V. G. A CTCF Code for 3D Genome Architecture. *Cell* **162,** 703–705 (2015).

73. Terakawa, T. *et al.* The condensin complex is a mechanochemical motor that translocates along DNA. *Science (80-.).* **676,** eaan6516 (2017).

74. Busslinger, G. A. *et al.* Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* **544,** 503–507 (2017).

75. Stevens, T. J. *et al.* 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544,** 1–21 (2017).

76. Andrey, G. & Mundlos, S. The three-dimensional genome: regulating gene expression during pluripotency and development. *Development* **144,** 3646–3658 (2017).

77. Beagan, J. A. *et al.* YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Res.* **27,** 1139–1152 (2017).

78. Bailey, S. D. *et al.* ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters. *Nat. Commun.* **2,** 6186 (2015).

79. Weintraub, A. S. *et al.* YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* **171,** 1573–1588.e28 (2017).

80. Phillips-Cremins, J. E. *et al.* Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. *Cell* **153,** 1281–1295 (2013).

81. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467,** 430–435 (2010).

82. Chen, X., Fang, F., Liou, Y.-C. & Ng, H.-H. Zfp143 Regulates Nanog Through Modulation of Oct4 Binding. *Stem Cells* **26,** 2759–2767 (2008).

83. Wei, Z. *et al.* Klf4 Organizes Long-Range Chromosomal Interactions with the Oct4 Locus in Reprogramming and Pluripotency. *Cell Stem Cell* **13,** 36–47 (2013).

84. Denholtz, M. *et al.* Long-Range Chromatin Contacts in Embryonic Stem Cells Reveal a Role for

Pluripotency Factors and Polycomb Proteins in Genome Organization. *Cell Stem Cell* **13,** 602–616 (2013).

85. van Steensel, B. & Belmont, A. S. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* **169,** 780–791 (2017).

86. Peric-Hupkes, D. *et al.* Molecular Maps of the Reorganization of Genome-Nuclear Lamina Interactions during Differentiation. *Mol. Cell* **38,** 603–613 (2010).

87. Therizols, P. *et al.* Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science (80-.).* **346,** 1238–1242 (2014).

88. Harr, J. C. *et al.* Directed targeting of chromatin to the nuclear lamina is mediated by chromatin state and A-type lamins. *J. Cell Biol.* **208,** 33–52 (2015).

89. Festuccia, N. *et al.* Mitotic binding of Esrrb marks key regulatory regions of the pluripotency network. *Nat. Cell Biol.* **18,** 1139–1148 (2016).

90. Liu, Y. *et al.* Widespread Mitotic Bookmarking by Histone Marks and Transcription Factors in Pluripotent Stem Cells. *Cell Rep.* **19,** 1283–1293 (2017).

91. Shao, Z. *et al.* Reprogramming by De-bookmarking the Somatic Transcriptional Program through Targeting of BET Bromodomains. *Cell Rep.* **16,** 3138–3145 (2016).

92. Teves, S. S. *et al.* A dynamic mode of mitotic bookmarking by transcription factors. *Elife* **5,** 1–24 (2016).

93. Palozola, K. C. *et al.* Mitotic transcription and waves of gene reactivation during mitotic exit. *Science (80-.).* **358,** 119–122 (2017).

94. Loh, K. M. & Lim, B. A precarious balance: Pluripotency factors as lineage specifiers. *Cell Stem Cell* **8,** 363–369 (2011).

95. Kopp, J. L., Ormsbee, B. D., Desler, M. & Rizzino, A. Small Increases in the Level of Sox2 Trigger the Differentiation of Mouse Embryonic Stem Cells. *Stem Cells* **26,** 903–911 (2008).

96. Yu, P., Pan, G., Yu, J. & Thomson, J. A. FGF2 Sustains NANOG and Switches the Outcome of BMP4-Induced Human Embryonic Stem Cell Differentiation. *Cell Stem Cell* **8,** 326–334 (2011).

97. Boyer, L. a *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122,** 947–56 (2005).

98. Tapia, N. *et al.* Dissecting the role of distinct OCT4-SOX2 heterodimer configurations in pluripotency. *Sci. Rep.* **5,** 13533 (2015).

99. Dailey, L. & Basilico, C. Coevolution of HMG domains and homeodomains and the generation of transcriptional regulation by Sox/POU complexes. *J. Cell. Physiol.* **186,** 315–328 (2001).

100. Aksoy, I. *et al.* Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. *EMBO J.* **32,** 938–953 (2013).

101. Merino, F. *et al.* Structural Basis for the SOX-Dependent Genomic Redistribution of OCT4 in Stem Cell Differentiation. *Structure* **22,** 1274–1286 (2014).

102. Gagliardi, A. *et al.* A direct physical interaction between Nanog and Sox2 regulates embryonic stem cell self-renewal. *EMBO J.* **32,** 2231–2247 (2013).

103. Wu, Q. *et al.* Sall4 Interacts with Nanog and Co-occupies Nanog Genomic Sites in Embryonic Stem Cells. *J. Biol. Chem.* **281,** 24090–24094 (2006).

104. van den Berg, D. L. C. *et al.* Estrogen-Related Receptor Beta Interacts with Oct4 To Positively Regulate Nanog Gene Expression. *Mol. Cell. Biol.* **28,** 5986–5995 (2008).

105. Costa, Y. *et al.* NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature* **495,** 370–374 (2013).

106. Yang, J., Gao, C., Chai, L. & Ma, Y. A Novel SALL4/OCT4 Transcriptional Feedback Network for Pluripotency of Embryonic Stem Cells. *PLoS One* **5,** e10766 (2010).

107. Chia, N.-Y. *et al.* A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature* **468,** 316–320 (2010).

108. Zhou, Q., Chipperfield, H., Melton, D. A. & Wong, W. H. A gene regulatory network in mouse embryonic stem cells. *Proc. Natl. Acad. Sci.* **104,** 16438–16443 (2007).

109. Kim, J., Chu, J., Shen, X., Wang, J. & Orkin, S. H. An Extended Transcriptional Network for Pluripotency of Embryonic Stem Cells. *Cell* **132,** 1049–1061 (2008).

110. Chen, X. *et al.* Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. *Cell* **133,** 1106–1117 (2008).

111. Rafiee, M. R., Girardot, C., Sigismondo, G. & Krijgsveld, J. Expanding the Circuitry of Pluripotency by Selective Isolation of Chromatin-Associated Proteins. *Mol. Cell* **64,** 624–635 (2016).

112. D'Alessio, A. C. *et al.* A Systematic Approach to Identify Candidate Transcription Factors that Control Cell Identity. *Stem Cell Reports* **5,** 763–775 (2015).

113. Cahan, P. *et al.* CellNet: Network Biology Applied to Stem Cell Engineering. *Cell* **158,** 903–915 (2014).

114. Heinäniemi, M. *et al.* Gene-pair expression signatures reveal lineage control. *Nat. Methods* **10,** 577–583 (2013).

115. Saint-André, V. *et al.* Models of human core transcriptional regulatory circuitries. *Genome Res.* **26,** 385–396 (2016).

116. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **17,** 487–500 (2016).

117. Li, E. & Zhang, Y. DNA Methylation in Mammals. *Cold Spring Harb. Perspect. Biol.* **6,** a019133–a019133 (2014).

118. Wu, T. P. *et al.* DNA methylation on N6-adenine in mammalian embryonic stem cells. *Nature* **532,** 329–333 (2016).

119. Fujita, N. *et al.* Methyl-CpG Binding Domain 1 (MBD1) Interacts with the Suv39h1-HP1 Heterochromatic Complex for DNA Methylation-based Transcriptional Repression. *J. Biol. Chem.* **278,** 24132–24138 (2003).

120. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21,** 381–395 (2011).

121. Black, J. C., Van Rechem, C. & Whetstine, J. R. Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact. *Mol. Cell* **48,** 491–507 (2012).

122. Marmorstein, R. & Zhou, M.-M. Writers and Readers of Histone Acetylation: Structure, Mechanism, and Inhibition. *Cold Spring Harb. Perspect. Biol.* **6,** a018762–a018762 (2014).

123. Ringrose, L. & Paro, R. Epigenetic Regulation of Cellular Memory by the Polycomb and Trithorax Group Proteins. *Annu. Rev. Genet.* **38,** 413–443 (2004).

124. Bernstein, B. E. *et al.* A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **125,** 315–326 (2006).

125. Azuara, V. *et al.* Chromatin signatures of pluripotent cell lines. *Nat. Cell Biol.* **8,** 532–538 (2006).

126. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9

creates a binding site for HP1 proteins. *Nature* **410,** 116–120 (2001).

127. Devaiah, B. N. *et al.* BRD4 is a histone acetyltransferase that evicts nucleosomes from chromatin. *Nat. Struct. Mol. Biol.* **23,** 540–548 (2016).

128. Itzen, F., Greifenberg, A. K., Bösken, C. a. & Geyer, M. Brd4 activates P-TEFb for RNA polymerase II CTD phosphorylation. *Nucleic Acids Res.* **42,** 7577–7590 (2014).

129. Yang, Z. *et al.* Recruitment of P-TEFb for Stimulation of Transcriptional Elongation by the Bromodomain Protein Brd4. *Mol. Cell* **19,** 535–545 (2005).

130. Jonkers, I. & Lis, J. T. Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16,** 167–177 (2015).

131. Chen, T. & Dent, S. Y. R. Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nat. Rev. Genet.* **15,** 93–106 (2014).

132. Gaspar-Maia, A., Alajem, A., Meshorer, E. & Ramalho-Santos, M. Open chromatin in pluripotency and reprogramming. *Nat. Rev. Mol. Cell Biol.* **12,** 36–47 (2011).

133. Meshorer, E. *et al.* Hyperdynamic Plasticity of Chromatin Proteins in Pluripotent Embryonic Stem Cells. *Dev. Cell* **10,** 105–116 (2006).

134. Efroni, S. *et al.* Global Transcription in Pluripotent Embryonic Stem Cells. *Cell Stem Cell* **2,** 437–447 (2008).

135. Novo, C. L. *et al.* The pluripotency factor Nanog regulates pericentromeric heterochromatin organization in mouse embryonic stem cells. *Genes Dev.* **30,** 1101–1115 (2016).

136. Gurdon, J. B. The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *J. Embryol. Exp. Morphol.* **10,** 622–40 (1962).

137. Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J. J. & Campbell, K. H. S. Viable offspring derived from fetal and adult mammalian cells. *Nature* **385,** 810–813 (1997).

138. Tada, M., Takahama, Y., Abe, K., Nakatsuji, N. & Tada, T. Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. *Curr. Biol.* **11,** 1553–1558 (2001).

139. Cowan, C. A., Atienza, J., Melton, D. A. & Eggan, K. Nuclear Reprogramming of Somatic Cells After Fusion with Human Embryonic Stem Cells. *Science (80-.).* **309,** 1369–1373 (2005).

140. Vierbuchen, T. *et al.* Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463,** 1035–1041 (2010).

141. Huang, P. *et al.* Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature* **475,** 386–389 (2011).

142. Sekiya, S. & Suzuki, A. Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. *Nature* **475,** 390–393 (2011).

143. Ieda, M. *et al.* Direct Reprogramming of Fibroblasts into Functional Cardiomyocytes by Defined Factors. *Cell* **142,** 375–386 (2010).

144. Sancho-Martinez, I., Baek, S. H. & Izpisua Belmonte, J. C. Lineage conversion methodologies meet the reprogramming toolbox. *Nat. Cell Biol.* **14,** 892–899 (2012).

145. Morris, S. A. *et al.* Dissecting Engineered Cell Types and Enhancing Cell Fate Conversion via CellNet. *Cell* **158,** 889–902 (2014).

146. Rackham, O. J. L. *et al.* A predictive computational framework for direct reprogramming between human cell types. *Nat. Genet.* 1–8 (2016). doi:10.1038/ng.3487

147. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem

cells. *Nature* **448,** 313–317 (2007).

148. Zhao, X. *et al.* iPS cells produce viable mice through tetraploid complementation. *Nature* **461,** 86–90 (2009).

149. Takahashi, K. *et al.* Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* **131,** 861–872 (2007).

150. Park, I.-H. *et al.* Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451,** 141–146 (2008).

151. Mullen, A. C. *et al.* Master Transcription Factors Determine Cell-Type-Specific Responses to TGF-β Signaling. *Cell* **147,** 565–576 (2011).

152. Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* **25,** 2227–2241 (2011).

153. Zaret, K. S., Caravaca, J. M., Tulin, A. & Sekiya, T. Nuclear Mobility and Mitotic Chromosome Binding: Similarities between Pioneer Transcription Factor FoxA and Linker Histone H1. *Cold Spring Harb. Symp. Quant. Biol.* **75,** 219–226 (2010).

154. Yan, J., Xu, L., Crawford, G., Wang, Z. & Burgess, S. M. The Forkhead Transcription Factor FoxI1 Remains Bound to Condensed Mitotic Chromosomes and Stably Remodels Chromatin Structure. *Mol. Cell. Biol.* **26,** 155–168 (2006).

155. Soufi, A. *et al.* Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161,** 555–568 (2015).

156. Wu, G. & Schöler, H. R. Role of Oct4 in the early embryo development. *Cell Regen.* **3,** 3:7 (2014).

157. Nichols, J. *et al.* Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4. *Cell* **95,** 379–391 (1998).

158. Niakan, K. K. & Eggan, K. Analysis of human embryos from zygote to blastocyst reveals distinct gene expression patterns relative to the mouse. *Dev. Biol.* **375,** 54–64 (2013).

159. Fogarty, N. M. E. *et al.* Genome editing reveals a role for OCT4 in human embryogenesis. *Nature* **550,** 67–73 (2017).

160. Wu, G. *et al.* Establishment of totipotency does not depend on Oct4A. *Nat. Cell Biol.* **15,** 1089–1097 (2013).

161. Zafarana, G., Avery, S. R., Avery, K., Moore, H. D. & Andrews, P. W. Specific Knockdown of OCT4 in Human Embryonic Stem Cells by Inducible Short Hairpin RNA Interference. *Stem Cells* **27,** 776–782 (2009).

162. Radzisheuskaya, A. & Silva, J. C. R. Do all roads lead to Oct4? The emerging concepts of induced pluripotency. *Trends Cell Biol.* **24,** 275–284 (2014).

163. Kim, J. B. *et al.* Oct4-Induced Pluripotency in Adult Neural Stem Cells. *Cell* **136,** 411–419 (2009).

164. Jerabek, S. *et al.* Changing POU dimerization preferences converts Oct6 into a pluripotency inducer. *EMBO Rep.* **18,** 319–333 (2017).

165. Thier, M. *et al.* Direct Conversion of Fibroblasts into Stably Expandable Neural Stem Cells. *Cell Stem Cell* **10,** 473–479 (2012).

166. Heng, J.-C. D. C. D. *et al.* The Nuclear Receptor Nr5a2 Can Replace Oct4 in the Reprogramming of Murine Somatic Cells to Pluripotent Cells. *Cell Stem Cell* **6,** 167–174 (2010).

167. Gao, Y. *et al.* Replacement of Oct4 by Tet1 during iPSC Induction Reveals an Important Role of DNA Methylation and Hydroxymethylation in Reprogramming. *Cell Stem Cell* **12,** 453–469 (2013).

168. Buganim, Y. *et al.* Single-Cell Expression Analyses during Cellular Reprogramming Reveal an Early Stochastic and a Late Hierarchic Phase. *Cell* **150,** 1209–1222 (2012).

169. Redmer, T. *et al.* E-cadherin is crucial for embryonic stem cell pluripotency and can replace OCT4 during somatic cell reprogramming. *EMBO Rep.* **12,** 720–726 (2011).

170. Fritz, A. L., Adil, M. M., Mao, S. R. & Schaffer, D. V. cAMP and EPAC Signaling Functionally Replace OCT4 During Induced Pluripotent Stem Cell Reprogramming. *Mol. Ther.* **23,** 952–963 (2015).

171. Montserrat, N. *et al.* Reprogramming of Human Fibroblasts to Pluripotency with Lineage Specifiers. *Cell Stem Cell* **13,** 1–10 (2013).

172. Shu, J. *et al.* Induction of Pluripotency in Mouse Somatic Cells with Lineage Specifiers. *Cell* **153,** 963–975 (2013).

173. Hammachi, F. *et al.* Transcriptional Activation by Oct4 Is Sufficient for the Maintenance and Induction of Pluripotency. *Cell Rep.* **1,** 99–109 (2012).

174. Hirai, H. *et al.* Radical Acceleration of Nuclear Reprogramming by Chromatin Remodeling With the Transactivation Domain of MyoD. *Stem Cells* **29,** N/A-N/A (2011).

175. Zhu, G. *et al.* Coordination of Engineered Factors with TET1/2 Promotes Early-Stage Epigenetic Modification during Somatic Cell Reprogramming. *Stem Cell Reports* **2,** 253–261 (2014).

176. Wang, Y. *et al.* Reprogramming of mouse and human somatic cells by high-performance engineered factors. *EMBO Rep.* **12,** 373–378 (2011).

177. Wang, L. *et al.* Enhanced human somatic cell reprogramming efficiency by fusion of the MYC transactivation domain and OCT4. *Stem Cell Res.* **25,** 88–97 (2017).

178. Keramari, M. *et al.* Sox2 Is Essential for Formation of Trophectoderm in the Preimplantation Embryo. *PLoS One* **5,** e13952 (2010).

179. Sarkar, A. & Hochedlinger, K. The Sox family of transcription factors: Versatile regulators of stem and progenitor cell fate. *Cell Stem Cell* **12,** 15–30 (2013).

180. Julian, L. M., McDonald, A. C. & Stanford, W. L. Direct reprogramming with SOX factors: masters of cell fate. *Curr. Opin. Genet. Dev.* **46,** 24–36 (2017).

181. Nakagawa, M. *et al.* Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat. Biotechnol.* **26,** 101–106 (2007).

182. Jauch, R. *et al.* Conversion of Sox17 into a Pluripotency Reprogramming Factor by Reengineering Its Association with Oct4 on DNA. *Stem Cells* **29,** 940–951 (2011).

183. Ring, K. L. *et al.* Direct Reprogramming of Mouse and Human Fibroblasts into Multipotent Neural Stem Cells with a Single Factor. *Cell Stem Cell* **11,** 100–109 (2012).

184. Sui, D. *et al.* Fine-Tuning of iPSC Derivation by an Inducible Reprogramming System at the Protein Level. *Stem Cell Reports* **2,** 721–733 (2014).

185. Bialkowska, A. B., Yang, V. W. & Mallipattu, S. K. Krüppel-like factors in mammalian stem cells and development. *Development* **144,** 737–754 (2017).

186. Blakeley, P. *et al.* Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development* **142,** 3151–3165 (2015).

187. Lin, S.-C. J., Wani, M. A., Whitsett, J. A. & Wells, J. M. Klf5 regulates lineage formation in the pre-implantation mouse embryo. *Development* **137,** 3953–3963 (2010).

188. Jiang, J. *et al.* A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nat. Cell Biol.* **10,** 353–360 (2008).

189. Jeon, H. *et al.* Comprehensive Identification of Krüppel-Like Factor Family Members Contributing to the Self-Renewal of Mouse Embryonic Stem Cells and Cellular Reprogramming. *PLoS One* **11,** e0150715 (2016).

190. Lyssiotis, C. a *et al.* Reprogramming of murine fibroblasts to induced pluripotent stem cells with chemical complementation of Klf4. *Proc. Natl. Acad. Sci.* **106,** 8912–8917 (2009).

191. Chen, J. *et al.* BMPs functionally replace Klf4 and support efficient reprogramming of mouse fibroblasts by Oct4 alone. *Cell Res.* **21,** 205–212 (2011).

192. Samavarchi-Tehrani, P. *et al.* Functional Genomics Reveals a BMP-Driven Mesenchymal-to-Epithelial Transition in the Initiation of Somatic Cell Reprogramming. *Cell Stem Cell* **7,** 64–77 (2010).

193. Li, R. *et al.* A Mesenchymal-to-Epithelial Transition Initiates and Is Required for the Nuclear Reprogramming of Mouse Fibroblasts. *Cell Stem Cell* **7,** 51–63 (2010).

194. Gumireddy, K. *et al.* KLF17 is a negative regulator of epithelial–mesenchymal transition and metastasis in breast cancer. *Nat. Cell Biol.* **11,** 1297–1304 (2009).

195. Nishimura, K. *et al.* Manipulation of KLF4 Expression Generates iPSCs Paused at Successive Stages of Reprogramming. *Stem Cell Reports* **3,** 915–929 (2014).

196. Koyanagi-Aoi, M. *et al.* Differentiation-defective phenotypes revealed by large-scale analyses of human pluripotent stem cells. *Proc. Natl. Acad. Sci.* **110,** 20569–20574 (2013).

197. Toivonen, S. *et al.* Comparative Analysis of Targeted Differentiation of Human Induced Pluripotent Stem Cells (hiPSCs) and Human Embryonic Stem Cells Reveals Variability Associated With Incomplete Transgene Silencing in Retrovirally Derived hiPSC Lines. *Stem Cells Transl. Med.* **2,** 83–93 (2013).

198. Wang, J. *et al.* Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* **516,** 405–409 (2014).

199. Ohnuki, M. *et al.* Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. *Proc. Natl. Acad. Sci.* **111,** 12426–12431 (2014).

200. Meyer, N. & Penn, L. Z. Reflecting on 25 years with MYC. *Nat. Rev. Cancer* **8,** 976–990 (2008).

201. Allevato, M. *et al.* Sequence-specific DNA binding by MYC/MAX to low-affinity non-E-box motifs. *PLoS One* **12,** e0180147 (2017).

202. Eberhardy, S. R. & Farnham, P. J. Myc recruits P-TEFb to mediate the final step in the transcriptional activation of the cad promoter. *J. Biol. Chem.* **277,** 40156–40162 (2002).

203. Rahl, P. B. *et al.* c-Myc Regulates Transcriptional Pause Release. *Cell* **141,** 432–445 (2010).

204. Nie, Z. *et al.* c-Myc Is a Universal Amplifier of Expressed Genes in Lymphocytes and Embryonic Stem Cells. *Cell* **151,** 68–79 (2012).

205. Lin, C. Y. Y. *et al.* Transcriptional Amplification in Tumor Cells with Elevated c-Myc. *Cell* **151,** 56–67 (2012).

206. Clavería, C., Giovinazzo, G., Sierra, R. & Torres, M. Myc-driven endogenous cell competition in the early mammalian embryo. *Nature* **500,** 39–44 (2013).

207. Scognamiglio, R. *et al.* Myc Depletion Induces a Pluripotent Dormant State Mimicking Diapause. *Cell* **164,** 668–680 (2016).

208. Bulut-Karslioglu, A. *et al.* Inhibition of mTOR induces a paused pluripotent state. *Nature* **540,** 119–123 (2016).

209. Kim, J. *et al.* A Myc Network Accounts for Similarities between Embryonic Stem and Cancer Cell Transcription Programs. *Cell* **143,** 313–324 (2010).

210. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and Impediments of the Pluripotency Reprogramming Factors' Initial Engagement with the Genome. *Cell* **151,** 994–1004 (2012).

211. Nichols, J. & Smith, A. Pluripotency in the Embryo and in Culture. *Cold Spring Harb. Perspect. Biol.* **4,** a008128–a008128 (2012).

212. Schwarz, B. A., Bar-Nur, O., Silva, J. C. R. & Hochedlinger, K. Nanog Is Dispensable for the Generation of Induced Pluripotent Stem Cells. *Curr. Biol.* **24,** 347–350 (2014).

213. Ebata, K. T. *et al.* Vitamin C induces specific demethylation of H3K9me2 in mouse embryonic stem cells via Kdm3a/b. *Epigenetics Chromatin* **10,** 36 (2017).

214. Hanna, J. *et al.* Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462,** 595–601 (2009).

215. Viswanathan, S. R. & Daley, G. Q. Lin28: A MicroRNA Regulator with a Macro Role. *Cell* **140,** 445–449 (2010).

216. Ruiz, S. *et al.* A High Proliferation Rate Is Required for Cell Reprogramming and Maintenance of Human Embryonic Stem Cell Identity. *Curr. Biol.* **21,** 45–52 (2011).

217. Lapasset, L. *et al.* Rejuvenating senescent and centenarian human cells by reprogramming through the pluripotent state. *Genes Dev.* **25,** 2248–2253 (2011).

218. Zhang, J. *et al.* LIN28 Regulates Stem Cell Metabolism and Conversion to Primed Pluripotency. *Cell Stem Cell* **19,** 66–80 (2016).

219. Zeng, Y. *et al.* Lin28A Binds Active Promoters and Recruits Tet1 to Regulate Gene Expression. *Mol. Cell* **61,** 153–160 (2016).

220. Liu, J. *et al.* The oncogene c-Jun impedes somatic cell reprogramming. *Nat. Cell Biol.* **17,** 856–867 (2015).

221. Mansour, A. A. *et al.* The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. *Nature* **488,** 409–413 (2012).

222. Moon, J.-H. *et al.* Reprogramming fibroblasts into induced pluripotent stem cells with Bmi1. *Cell Res.* **21,** 1305–1315 (2011).

223. Gonzalez-Munoz, E., Arboleda-Estudillo, Y., Otu, H. H. & Cibelli, J. B. Histone chaperone ASF1A is required for maintenance of pluripotency and cellular reprogramming. *Science (80-.).* **345,** 822–825 (2014).

224. Chen, J. *et al.* The Combination of Tet1 with Oct4 Generates High-Quality Mouse-Induced Pluripotent Stem Cells. *Stem Cells* **33,** 686–698 (2015).

225. Anokye-Danso, F. *et al.* Highly Efficient miRNA-Mediated Reprogramming of Mouse and Human Somatic Cells to Pluripotency. *Cell Stem Cell* **8,** 376–388 (2011).

226. Miyoshi, N. *et al.* Reprogramming of Mouse and Human Cells to Pluripotency Using Mature MicroRNAs. *Cell Stem Cell* **8,** 633–638 (2011).

227. Smith, Z. D., Nachman, I., Regev, A. & Meissner, A. Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. *Nat. Biotechnol.* **28,** 521–526 (2010).

228. Polo, J. M. *et al.* A Molecular Roadmap of Reprogramming Somatic Cells into iPS Cells. *Cell* **151,** 1617–1632 (2012).

229. Taberlay, P. C. *et al.* Polycomb-Repressed Genes Have Permissive Enhancers that Initiate Reprogramming. *Cell* **147,** 1283–1294 (2011).

230. You, J. S. *et al.* OCT4 establishes and maintains nucleosome-depleted regions that provide additional

layers of epigenetic regulation of its target genes. *Proc. Natl. Acad. Sci.* **108,** 14497–14502 (2011).

231. He, S. *et al.* Passive DNA demethylation preferentially up-regulates pluripotency-related genes and facilitates the generation of induced pluripotent stem cells. *J. Biol. Chem.* **292,** 18542–18555 (2017).

232. Papp, B. & Plath, K. Epigenetics of reprogramming to induced pluripotency. *Cell* **152,** 1324–1343 (2013).

233. Chronis, C. *et al.* Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell* **168,** 442–459.e20 (2017).

234. Liu, X. *et al.* Sequential introduction of reprogramming factors reveals a time-sensitive requirement for individual factors and a sequential EMT–MET mechanism for optimal reprogramming. *Nat. Cell Biol.* **15,** 829–838 (2013).

235. Onder, T. T. *et al.* Chromatin-modifying enzymes as modulators of reprogramming. *Nature* **483,** 598–602 (2012).

236. Banito, A. & Gil, J. Induced pluripotent stem cells and senescence: learning the biology to improve the technology. *EMBO Rep.* **11,** 353–359 (2010).

237. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454,** 49–55 (2008).

238. Hong, H. *et al.* Suppression of induced pluripotent stem cell generation by the p53–p21 pathway. *Nature* **460,** 1132–1135 (2009).

239. Kawamura, T. *et al.* Linking the p53 tumour suppressor pathway to somatic cell reprogramming. *Nature* **460,** 1140–1144 (2009).

240. Marión, R. M. *et al.* A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. *Nature* **460,** 1149–1153 (2009).

241. Li, H. *et al.* The Ink4/Arf locus is a barrier for iPS cell reprogramming. *Nature* **460,** 1136–1139 (2009).

242. Banito, A. *et al.* Senescence impairs successful reprogramming to pluripotent stem cells. *Genes Dev.* **23,** 2134–2139 (2009).

243. Li, Y. *et al.* The p53–PUMA axis suppresses iPSC generation. *Nat. Commun.* **4,** 2174 (2013).

244. Hussein, S. M. I., Elbaz, J. & Nagy, A. A. Genome damage in induced pluripotent stem cells: Assessing the mechanisms and their consequences. *BioEssays* **35,** 152–162 (2013).

245. Ruiz, S. *et al.* Limiting replication stress during somatic cell reprogramming reduces genomic instability in induced pluripotent stem cells. *Nat. Commun.* **6,** 8036 (2015).

246. Son, M. J. *et al.* Nicotinamide Overcomes Pluripotency Deficits and Reprogramming Barriers. *Stem Cells* **31,** 1121–1135 (2013).

247. Ji, J. *et al.* Antioxidant Supplementation Reduces Genomic Aberrations in Human Induced Pluripotent Stem Cells. *Stem Cell Reports* **2,** 44–51 (2014).

248. Ji, S. *et al.* Baf60b-mediated ATM-p53 activation blocks cell identity conversion by sensing chromatin opening. *Cell Res.* **27,** 642–656 (2017).

249. Bunch, H. *et al.* Transcriptional elongation requires DNA break-induced signalling. *Nat. Commun.* **6,** 10191 (2015).

250. Wu, J., Ocampo, A. & Belmonte, J. C. I. Cellular Metabolism and Induced Pluripotency. *Cell* **166,** 1371–1385 (2016).

251. Kida, Y. S. *et al.* ERRs Mediate a Metabolic Switch Required for Somatic Cell Reprogramming to Pluripotency. *Cell Stem Cell* **16,** 547–555 (2015).

252. Hawkins, K. E. *et al.* NRF2 Orchestrates the Metabolic Shift during Induced Pluripotent Stem Cell Reprogramming. *Cell Rep.* **14,** 1883–1891 (2016).

253. Mathieu, J. *et al.* Hypoxia-Inducible Factors Have Distinct and Stage-Specific Roles during Reprogramming of Human Cells to Pluripotency. *Cell Stem Cell* **14,** 592–605 (2014).

254. Yoshida, Y., Takahashi, K., Okita, K., Ichisaka, T. & Yamanaka, S. Hypoxia Enhances the Generation of Induced Pluripotent Stem Cells. *Cell Stem Cell* **5,** 237–241 (2009).

255. Zhu, S. *et al.* Reprogramming of Human Primary Somatic Cells by OCT4 and Chemical Compounds. *Cell Stem Cell* **7,** 651–655 (2010).

256. Khaw, S.-L., Min-Wen, C., Koh, C.-G., Lim, B. & Shyh-Chang, N. Oocyte Factors Suppress Mitochondrial Polynucleotide Phosphorylase to Remodel the Metabolome and Enhance Reprogramming. *Cell Rep.* **12,** 1080–1088 (2015).

257. Folmes, C. D. L. *et al.* Somatic Oxidative Bioenergetics Transitions into Pluripotency-Dependent Glycolysis to Facilitate Nuclear Reprogramming. *Cell Metab.* **14,** 264–271 (2011).

258. Chen, J. *et al.* Hierarchical Oct4 Binding in Concert with Primed Epigenetic Rearrangements during Somatic Cell Reprogramming. *Cell Rep.* **14,** 1540–1554 (2016).

259. Hussein, S. M. I. *et al.* Genome-wide characterization of the routes to pluripotency. *Nature* **516,** 198–206 (2014).

260. Xu, Y. *et al.* Transcriptional Control of Somatic Cell Reprogramming. *Trends Cell Biol.* **xx,** 1–17 (2016).

261. Lee, D.-S. *et al.* An epigenomic roadmap to induced pluripotency reveals DNA methylation as a reprogramming modulator. *Nat. Commun.* **5,** 5619 (2014).

262. Huangfu, D. *et al.* Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nat. Biotechnol.* **26,** 795–797 (2008).

263. Doege, C. A. *et al.* Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. *Nature* **488,** 652–655 (2012).

264. Chen, J. *et al.* H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nat. Genet.* **45,** 34–42 (2012).

265. Sridharan, R. *et al.* Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1γ in reprogramming to pluripotency. *Nat. Cell Biol.* **15,** 872–882 (2013).

266. Miles, D. C. *et al.* TRIM28 Is an Epigenetic Barrier to Induced Pluripotent Stem Cell Reprogramming. *Stem Cells* **35,** 147–157 (2017).

267. Cheloufi, S. *et al.* The histone chaperone CAF-1 safeguards somatic cell identity. *Nature* **528,** 218–224 (2015).

268. Matoba, S. *et al.* Embryonic Development following Somatic Cell Nuclear Transfer Impeded by Persisting Histone Methylation. *Cell* **159,** 884–895 (2014).

269. Rais, Y. *et al.* Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502,** 65–70 (2013).

270. Yildirim, O. *et al.* Mbd3/NURD Complex Regulates Expression of 5-Hydroxymethylcytosine Marked Genes in Embryonic Stem Cells. *Cell* **147,** 1498–1510 (2011).

271. Spruijt, C. G. *et al.* Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives. *Cell* **152,** 1146–1159 (2013).

272. Hashimoto, H. *et al.* Recognition and potential mechanisms for replication and erasure of cytosine

hydroxymethylation. *Nucleic Acids Res.* **40,** 4841–4849 (2012).

273. dos Santos, R. L. *et al.* MBD3/NuRD Facilitates Induction of Pluripotency in a Context-Dependent Manner. *Cell Stem Cell* **15,** 102–110 (2014).

274. Di Stefano, B. *et al.* C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells. *Nature* **506,** 235–239 (2013).

275. Di Stefano, B. *et al.* C/EBPα creates elite cells for iPSC reprogramming by upregulating Klf4 and increasing the levels of Lsd1 and Brd4. *Nat. Cell Biol.* **18,** 371–381 (2016).

276. O'Malley, J. *et al.* High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. *Nature* **499,** 88–91 (2013).

277. Takahashi, K. *et al.* Induction of pluripotency in human somatic cells via a transient state resembling primitive streak-like mesendoderm. *Nat. Commun.* **5,** 3678 (2014).

278. Cacchiarelli, D. *et al.* Integrative Analyses of Human Reprogramming Reveal Dynamic Nature of Induced Pluripotency. *Cell* **162,** 412–424 (2015).

279. Efe, J. a. *et al.* Conversion of mouse fibroblasts into cardiomyocytes using a direct reprogramming strategy. *Nat. Cell Biol.* **13,** 215–222 (2011).

280. Golipour, A. *et al.* A Late Transition in Somatic Cell Reprogramming Requires Regulators Distinct from the Pluripotency Network. *Cell Stem Cell* **11,** 769–782 (2012).

281. Amlani, B. *et al.* Nascent Induced Pluripotent Stem Cells Efficiently Generate Entirely iPSC-Derived Mice while Expressing Differentiation-Associated Genes. *Cell Rep.* **22,** 876–884 (2018).

282. Ebrahimi, B. Reprogramming barriers and enhancers: strategies to enhance the efficiency and kinetics of induced pluripotency. *Cell Regen.* **4,** 4:10 (2015).

283. Borkent, M. *et al.* A Serial shRNA Screen for Roadblocks to Reprogramming Identifies the Protein Modifier SUMO2. *Stem Cell Reports* **6,** 704–716 (2016).

284. Michlits, G. *et al.* CRISPR-UMI: single-cell lineage tracing of pooled CRISPR–Cas9 screens. *Nat. Methods* (2017). doi:10.1038/nmeth.4466

285. Qin, H. *et al.* Systematic Identification of Barriers to Human iPSC Generation. *Cell* **158,** 449–461 (2014).

286. Buckley, S. M. *et al.* Regulation of Pluripotency and Cellular Reprogramming by the Ubiquitin-Proteasome System. *Cell Stem Cell* **11,** 783–798 (2012).

287. Cai, N., Li, M., Qu, J., Liu, G.-H. & Izpisua Belmonte, J. C. Post-translational modulation of pluripotency. *J. Mol. Cell Biol.* **4,** 262–265 (2012).

288. Zhou, Z. *et al.* Kdm2b Regulates Somatic Reprogramming through Variant PRC1 Complex-Dependent Function. *Cell Rep.* **21,** 2160–2170 (2017).

289. Wong, M., Polly, P. & Liu, T. The histone methyltransferase DOT1L: regulatory functions and a cancer therapy target. *Am. J. Cancer Res.* **5,** 2823–37 (2015).

290. Koche, R. P. *et al.* Reprogramming Factor Expression Initiates Widespread Targeted Chromatin Remodeling. *Cell Stem Cell* **8,** 96–105 (2011).

291. Ho, L. & Crabtree, G. R. Chromatin remodelling during development. *Nature* **463,** 474–484 (2010).

292. Pardo, M. *et al.* An Expanded Oct4 Interaction Network: Implications for Stem Cell Biology, Development, and Disease. *Cell Stem Cell* **6,** 382–395 (2010).

293. van den Berg, D. L. C. *et al.* An Oct4-Centered Protein Interaction Network in Embryonic Stem Cells. *Cell Stem Cell* **6,** 369–381 (2010).

294. Esch, D. *et al.* A unique Oct4 interface is crucial for reprogramming to pluripotency. *Nat. Cell Biol.* **15,** 295–301 (2013).

295. Beneke, S. Regulation of chromatin structure by poly(ADP-ribosyl)ation. *Front. Genet.* **3,** 1–16 (2012).

296. Ruetz, T. *et al.* Constitutively Active SMAD2/3 Are Broad-Scope Potentiators of Transcription-Factor-Mediated Cellular Reprogramming. *Cell Stem Cell* 1–15 (2017). doi:10.1016/j.stem.2017.10.013

297. Polo, J. M. *et al.* Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nat. Biotechnol.* **28,** 848–855 (2010).

298. Lister, R. *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471,** 68–73 (2011).

299. Wang, T. *et al.* Subtelomeric hotspots of aberrant 5-hydroxymethylcytosine-mediated epigenetic modifications during reprogramming to pluripotency. *Nat. Cell Biol.* **15,** 700–711 (2013).

300. Huang, Y. *et al.* The Behaviour of 5-Hydroxymethylcytosine in Bisulfite Sequencing. *PLoS One* **5,** e8888 (2010).

301. Bhutani, N. *et al.* Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* **463,** 1042–1047 (2010).

302. Kumar, R. *et al.* AID stabilizes stem-cell phenotype by removing epigenetic memory of pluripotency genes. *Nature* **500,** 89–92 (2013).

303. Rangam, G., Schmitz, K.-M., Cobb, A. J. A. & Petersen-Mahrt, S. K. AID Enzymatic Activity Is Inversely Proportional to the Size of Cytosine C5 Orbital Cloud. *PLoS One* **7,** e43279 (2012).

304. Nabel, C. S. *et al.* AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol.* **8,** 751–758 (2012).

305. Planello, A. C. *et al.* Aberrant DNA methylation reprogramming during induced pluripotent stem cell generation is dependent on the choice of reprogramming factors. *Cell Regen.* **3,** 3:4 (2014).

306. Kohli, R. M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502,** 472–479 (2013).

307. Maiti, A. & Drohat, A. C. Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine. *J. Biol. Chem.* **286,** 35334–35338 (2011).

308. He, Y.-F. *et al.* Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science (80-.).* **333,** 1303–1307 (2011).

309. Bar, S., Schachter, M., Eldar-Geva, T. & Benvenisty, N. Large-Scale Analysis of Loss of Imprinting in Human Pluripotent Stem Cells. *Cell Rep.* **19,** 957–968 (2017).

310. Stadtfeld, M. *et al.* Ascorbic acid prevents loss of Dlk1-Dio3 imprinting and facilitates generation of all–iPS cell mice from terminally differentiated B cells. *Nat. Genet.* **44,** 398–405 (2012).

311. Young, J. I., Züchner, S. & Wang, G. Regulation of the Epigenome by Vitamin C. *Annu. Rev. Nutr.* **35,** 545–564 (2015).

312. Tomoda, K. *et al.* Derivation Conditions Impact X-Inactivation Status in Female Human Induced Pluripotent Stem Cells. *Cell Stem Cell* **11,** 91–99 (2012).

313. Mekhoubad, S. *et al.* Erosion of Dosage Compensation Impacts Human iPSC Disease Modeling. *Cell Stem Cell* **10,** 595–609 (2012).

314. Krijger, P. H. L. *et al.* Cell-of-Origin-Specific 3D Genome Structure Acquired during Somatic Cell Reprogramming. *Cell Stem Cell* **18,** 597–610 (2016).

315. Beagan, J. A. *et al.* Local Genome Topology Can Exhibit an Incompletely Rewired 3D-Folding State during Somatic Cell Reprogramming. *Cell Stem Cell* **18,** 611–624 (2016).

316. Leitch, H. G. *et al.* Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* **20,** 311–316 (2013).

317. Rodriguez, C. *et al.* CTCF is a DNA methylation-sensitive positive regulator of the INK/ARF locus. *Biochem. Biophys. Res. Commun.* **392,** 129–134 (2010).

318. Stadhouders, R. *et al.* Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. *Nat. Genet.* (2018). doi:10.1038/s41588-017-0030-7

319. Li, W., Li, K., Wei, W. & Ding, S. Chemical Approaches to Stem Cell Biology and Therapeutics. *Cell Stem Cell* **13,** 270–283 (2013).

320. Liang, G., Taranova, O., Xia, K. & Zhang, Y. Butyrate Promotes Induced Pluripotent Stem Cell Generation. *J. Biol. Chem.* **285,** 25516–25521 (2010).

321. Mali, P. *et al.* Butyrate Greatly Enhances Derivation of Human Induced Pluripotent Stem Cells by Promoting Epigenetic Remodeling and the Expression of Pluripotency-Associated Genes. *Stem Cells* **28,** 713–720 (2010).

322. Shi, Y. *et al.* A Combined Chemical and Genetic Approach for the Generation of Induced Pluripotent Stem Cells. *Cell Stem Cell* **2,** 525–528 (2008).

323. Shi, Y. *et al.* Induction of Pluripotent Stem Cells from Mouse Embryonic Fibroblasts by Oct4 and Klf4 with Small-Molecule Compounds. *Cell Stem Cell* **3,** 568–574 (2008).

324. Li, W. *et al.* Generation of Human Induced Pluripotent Stem Cells in the Absence of Exogenous Sox2. *Stem Cells* **27,** N/A-N/A (2009).

325. Esteban, M. A. *et al.* Vitamin C Enhances the Generation of Mouse and Human Induced Pluripotent Stem Cells. *Cell Stem Cell* **6,** 71–79 (2010).

326. Blaschke, K. *et al.* Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* **500,** 222–226 (2013).

327. Wang, T. *et al.* The Histone Demethylases Jhdm1a/1b Enhance Somatic Cell Reprogramming in a Vitamin-C-Dependent Manner. *Cell Stem Cell* **9,** 575–587 (2011).

328. Yuan, X. *et al.* Brief Report: Combined Chemical Treatment Enables Oct4-Induced Reprogramming from Mouse Embryonic Fibroblasts. *Stem Cells* **29,** 549–553 (2011).

329. Ichida, J. K. *et al.* A Small-Molecule Inhibitor of Tgf-β Signaling Replaces Sox2 in Reprogramming by Inducing Nanog. *Cell Stem Cell* **5,** 491–503 (2009).

330. Lin, T. *et al.* A chemical platform for improved induction of human iPSCs. *Nat. Methods* **6,** 805–808 (2009).

331. Wang, Y. & Adjaye, J. A Cyclic AMP Analog, 8-Br-cAMP, Enhances the Induction of Pluripotency in Human Fibroblast Cells. *Stem Cell Rev. Reports* **7,** 331–341 (2011).

332. Staerk, J. *et al.* Pan-Src Family Kinase Inhibitors Replace Sox2 during the Direct Reprogramming of Somatic Cells. *Angew. Chemie Int. Ed.* **50,** 5734–5736 (2011).

333. Li, W. *et al.* Identification of Oct4-activating compounds that enhance reprogramming efficiency. *Proc. Natl. Acad. Sci.* **109,** 20853–20858 (2012).

334. Li, Y. *et al.* Generation of iPSCs from mouse fibroblasts with a single gene, Oct4 and small molecules. *Cell Res.* **21,** 196–204 (2011).

335. Hou, P. *et al.* Pluripotent Stem Cells Induced from Mouse Somatic Cells by Small-Molecule Compounds. *Science (80-.).* **341,** 651–654 (2013).

336. Xie, X., Fu, Y. & Liu, J. Chemical reprogramming and transdifferentiation. *Curr. Opin. Genet. Dev.* **46,** 104–113 (2017).

337. Kuno, A., Nishimura, K. & Takahashi, S. Time-course transcriptome analysis of human cellular reprogramming from multiple cell types reveals the drastic change occurs between the mid phase and the late phase. *BMC Genomics* **19,** 9 (2018).

338. Knaupp, A. S. *et al.* Transient and Permanent Reconfiguration of Chromatin and Transcription Factor Occupancy Drive Reprogramming. *Cell Stem Cell* **21,** 834–845.e6 (2017).

339. Li, D. *et al.* Chromatin Accessibility Dynamics during iPSC Reprogramming. *Cell Stem Cell* **21,** 819–833.e6 (2017).

340. Nefzger, C. M. *et al.* Cell Type of Origin Dictates the Route to Pluripotency. *Cell Rep.* **21,** 2649–2660 (2017).

341. Zhao, Y. *et al.* A XEN-like State Bridges Somatic Cells to Pluripotency during Chemical Reprogramming. *Cell* **163,** 1678–1691 (2015).

342. Raab, S. *et al.* Reprogramming to pluripotency does not require transition through a primitive streak-like state. *Sci. Rep.* **7,** 16543 (2017).

343. Becker, J. S., Nicetto, D. & Zaret, K. S. H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. *Trends Genet.* **32,** 29–41 (2016).

344. Hu, K. Vectorology and Factor Delivery in Induced Pluripotent Stem Cell Reprogramming. *Stem Cells Dev.* **23,** 1301–1315 (2014).

345. Kitamura, T. *et al.* Retrovirus-mediated gene transfer and expression cloning: powerful tools in functional genomics. *Exp. Hematol.* **31,** 1007–1014 (2003).

346. Escors, D. & Breckpot, K. Lentiviral Vectors in Gene Therapy: Their Current Status and Future Potential. *Arch. Immunol. Ther. Exp. (Warsz).* **58,** 107–119 (2010).

347. Yang, B. X. *et al.* Systematic Identification of Factors for Provirus Silencing in Embryonic Stem Cells. *Cell* **163,** 230–245 (2015).

348. Maherali, N. *et al.* A High-Efficiency System for the Generation and Study of Human Induced Pluripotent Stem Cells. *Cell Stem Cell* **3,** 340–345 (2008).

349. Kane, N. M. *et al.* Lentivirus-mediated Reprogramming of Somatic Cells in the Absence of Transgenic Transcription Factors. *Mol. Ther.* **18,** 2139–2145 (2010).

350. Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T. & Yamanaka, S. Generation of Mouse Induced Pluripotent Stem Cells Without Viral Vectors. *Science (80-.).* **322,** 949–953 (2008).

351. Jia, F. *et al.* A nonviral minicircle vector for deriving human iPS cells. *Nat. Methods* **7,** 197–199 (2010).

352. Yu, J. *et al.* Human Induced Pluripotent Stem Cells Free of Vector and Transgene Sequences. *Science (80-.).* **324,** 797–801 (2009).

353. Okita, K. *et al.* A more efficient method to generate integration-free human iPS cells. *Nat. Methods* **8,** 409–412 (2011).

354. Wang, W. *et al.* Chromosomal transposition of PiggyBac in mouse embryonic stem cells. *Proc. Natl. Acad. Sci.* **105,** 9290–9295 (2008).

355. Cary, L. C. *et al.* Transposon mutagenesis of baculoviruses: Analysis of Trichoplusia ni transposon

IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. *Virology* **172,** 156–169 (1989).

356. Woltjen, K. *et al.* piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* **458,** 766–770 (2009).

357. Kaji, K. *et al.* Virus-free induction of pluripotency and subsequent excision of reprogramming factors. *Nature* **458,** 771–775 (2009).

358. Ivics, Z., Hackett, P. B., Plasterk, R. H. & Izsvák, Z. Molecular Reconstruction of Sleeping Beauty, a Tc1-like Transposon from Fish, and Its Transposition in Human Cells. *Cell* **91,** 501–510 (1997).

359. Grabundzija, I. *et al.* Sleeping Beauty transposon-based system for cellular reprogramming and targeted gene insertion in induced pluripotent stem cells. *Nucleic Acids Res.* **41,** 1829–1847 (2013).

360. Yusa, K., Rad, R., Takeda, J. & Bradley, A. Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon. *Nat. Methods* **6,** 363–369 (2009).

361. Warren, L. *et al.* Highly Efficient Reprogramming to Pluripotency and Directed Differentiation of Human Cells with Synthetic Modified mRNA. *Cell Stem Cell* **7,** 618–630 (2010).

362. Seth, R. B., Sun, L. & Chen, Z. J. Antiviral innate immunity pathways. *Cell Res.* **16,** 141–147 (2006).

363. Kell, A. M. & Gale, M. RIG-I in RNA virus recognition. *Virology* **479–480,** 110–121 (2015).

364. Nallagatla, S. R., Toroney, R. & Bevilacqua, P. C. A brilliant disguise for self RNA: 5'-end and internal modifications of primary transcripts suppress elements of innate immunity. *RNA Biol.* **5,** 140–144 (2008).

365. Mayo, C. B. & Cole, J. L. Interaction of PKR with single-stranded RNA. *Sci. Rep.* **7,** 3335 (2017).

366. Angel, M. & Yanik, M. F. Innate Immune Suppression Enables Frequent Transfection with RNA Encoding Reprogramming Proteins. *PLoS One* **5,** e11756 (2010).

367. Poleganov, M. A. *et al.* Efficient Reprogramming of Human Fibroblasts and Blood-Derived Endothelial Progenitor Cells Using Nonmodified RNA for Reprogramming and Immune Evasion. *Hum. Gene Ther.* **26,** 751–766 (2015).

368. Schlaeger, T. M. *et al.* A comparison of non-integrating reprogramming methods. *Nat. Biotechnol.* **33,** 58–63 (2014).

369. Lin, S.-L. *et al.* Mir-302 reprograms human skin cancer cells into a pluripotent ES-cell-like state. *RNA* **14,** 2115–2124 (2008).

370. Subramanyam, D. *et al.* Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells. *Nat. Biotechnol.* **29,** 443–448 (2011).

371. Lee, Y. J. *et al.* Dissecting microRNA-mediated regulation of stemness, reprogramming, and pluripotency. *Cell Regen.* **5,** 5:2 (2016).

372. Judson, R. L., Babiarz, J. E., Venere, M. & Blelloch, R. Embryonic stem cell–specific microRNAs promote induced pluripotency. *Nat. Biotechnol.* **27,** 459–461 (2009).

373. Zhang, Z. *et al.* Dissecting the Roles of miR-302/367 Cluster in Cellular Reprogramming Using TALE-based Repressor and TALEN. *Stem Cell Reports* **1,** 218–225 (2013).

374. Strauss, J. H. & Strauss, E. G. The alphaviruses: gene expression, replication, and evolution. *Microbiol. Rev.* **58,** 491–562 (1994).

375. Frolov, I., Garmashova, N., Atasheva, S. & Frolova, E. I. Random Insertion Mutagenesis of Sindbis Virus Nonstructural Protein 2 and Selection of Variants Incapable of Downregulating Cellular Transcription. *J. Virol.* **83,** 9031–9044 (2009).

376. Frolov, I. *et al.* Selection of RNA replicons capable of persistent noncytopathic replication in

mammalian cells. *J. Virol.* **73,** 3854–65 (1999).

377. Casales, E. *et al.* Development of a new noncytopathic Semliki Forest virus vector providing high expression levels and stability. *Virology* **376,** 242–251 (2008).

378. Pohjala, L. *et al.* Inhibitors of alphavirus entry and replication identified with a stable Chikungunya replicon cell line and virus-based assays. *PLoS One* **6,** e28923 (2011).

379. Petrakova, O. *et al.* Noncytopathic Replication of Venezuelan Equine Encephalitis Virus and Eastern Equine Encephalitis Virus Replicons in Mammalian Cells. *J. Virol.* **79,** 7597–7608 (2005).

380. Baccala, R., Hoebe, K., Kono, D. H., Beutler, B. & Theofilopoulos, A. N. TLR-dependent and TLR-independent pathways of type I interferon induction in systemic autoimmunity. *Nat. Med.* **13,** 543–551 (2007).

381. Kawai, T. *et al.* IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. *Nat. Immunol.* **6,** 981–988 (2005).

382. Burke, C. W., Gardner, C. L., Steffan, J. J., Ryman, K. D. & Klimstra, W. B. Characteristics of alpha/beta interferon induction after infection of murine fibroblasts with wild-type and mutant alphaviruses. *Virology* **395,** 121–132 (2009).

383. Lei, Y. *et al.* MAVS-Mediated Apoptosis and Its Inhibition by Viral Proteins. *PLoS One* **4,** e5466 (2009).

384. White, L. K. *et al.* Chikungunya Virus Induces IPS-1-Dependent Innate Immune Activation and Protein Kinase R-Independent Translational Shutoff. *J. Virol.* **85,** 606–620 (2011).

385. El Maadidi, S. *et al.* A Novel Mitochondrial MAVS/Caspase-8 Platform Links RNA Virus-Induced Innate Antiviral Signaling to Bax/Bak-Independent Apoptosis. *J. Immunol.* **192,** 1171–1183 (2014).

386. Yoshioka, N. *et al.* Efficient Generation of Human iPSCs by a Synthetic Self-Replicative RNA. *Cell Stem Cell* **13,** 246–254 (2013).

387. Guidotti, G., Brambilla, L. & Rossi, D. Cell-Penetrating Peptides: From Basic Research to Clinics. *Trends Pharmacol. Sci.* **38,** 406–424 (2017).

388. Lo, S. L. & Wang, S. An endosomolytic Tat peptide produced by incorporation of histidine and cysteine residues as a nonviral vector for DNA transfection. *Biomaterials* **29,** 2408–2414 (2008).

389. Pan, C., Lu, B., Chen, H. & Bishop, C. E. Reprogramming human fibroblasts using HIV-1 TAT recombinant proteins OCT4, SOX2, KLF4 and c-MYC. *Mol. Biol. Rep.* **37,** 2117–2124 (2010).

390. Kim, D. *et al.* Generation of Human Induced Pluripotent Stem Cells by Direct Delivery of Reprogramming Proteins. *Cell Stem Cell* **4,** 472–476 (2009).

391. Zhou, H. *et al.* Generation of Induced Pluripotent Stem Cells Using Recombinant Proteins. *Cell Stem Cell* **4,** 381–384 (2009).

392. Lee, J. *et al.* Activation of Innate Immunity Is Required for Efficient Nuclear Reprogramming. *Cell* **151,** 547–558 (2012).

393. Cho, H.-J. *et al.* Induction of pluripotent stem cells from adult somatic cells by protein-based reprogramming without genetic manipulation. *Blood* **116,** 386–395 (2010).

394. Stadtfeld, M., Nagaya, M., Utikal, J., Weir, G. & Hochedlinger, K. Induced Pluripotent Stem Cells Generated Without Viral Integration. *Science (80-.).* **322,** 945–949 (2008).

395. Li, H.-O. *et al.* A Cytoplasmic RNA Vector Derived from Nontransmissible Sendai Virus with Efficient Gene Transfer and Expression. *J. Virol.* **74,** 6564–6569 (2000).

396. Fusaki, N., Ban, H., Nishiyama, A., Saeki, K. & Hasegawa, M. Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not

integrate into the host genome. *Proc. Jpn. Acad. Ser. B. Phys. Biol. Sci.* **85,** 348–62 (2009).

397. Ban, H. *et al.* Efficient generation of transgene-free human induced pluripotent stem cells (iPSCs) by temperature-sensitive Sendai virus vectors. *Proc. Natl. Acad. Sci.* **108,** 14234–14239 (2011).

398. Nishimura, K. *et al.* Development of Defective and Persistent Sendai Virus Vector. *J. Biol. Chem.* **286,** 4760–4771 (2011).

399. Nishimura, K. *et al.* Simple and effective generation of transgene-free induced pluripotent stem cells using an auto-erasable Sendai virus vector responding to microRNA-302. *Stem Cell Res.* **23,** 13–19 (2017).

400. Trokovic, R. *et al.* Advanced Feeder-Free Generation of Induced Pluripotent Stem Cells Directly From Blood Cells. *Stem Cells Transl. Med.* **3,** 1402–1409 (2014).

401. Atchison, R. W., Casto, B. C. & Hammon, W. M. Adenovirus-Associated Defective Virus Particles. *Science (80-.).* **149,** 754–755 (1965).

402. Carter, B. J. Adeno-associated virus and the development of adeno-associated virus vectors: a historical perspective. *Mol. Ther.* **10,** 981–989 (2004).

403. Im, D. S. & Muzyczka, N. Partial purification of adeno-associated virus Rep78, Rep52, and Rep40 and their biochemical characterization. *J. Virol.* **66,** 1119–28 (1992).

404. Daya, S. & Berns, K. I. Gene Therapy Using Adeno-Associated Virus Vectors. *Clin. Microbiol. Rev.* **21,** 583–593 (2008).

405. Pereira, D. J., McCarty, D. M. & Muzyczka, N. The adeno-associated virus (AAV) Rep protein acts as both a repressor and an activator to regulate AAV transcription during a productive infection. *J. Virol.* **71,** 1079–88 (1997).

406. Kotin, R. M., Linden, R. M. & Berns, K. I. Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. *EMBO J.* **11,** 5071–8 (1992).

407. Weitzman, M. D., Kyostio, S. R., Kotin, R. M. & Owens, R. A. Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. *Proc. Natl. Acad. Sci.* **91,** 5808–5812 (1994).

408. Cataldi, M. P. & McCarty, D. M. Differential Effects of DNA Double-Strand Break Repair Pathways on Single-Strand and Self-Complementary Adeno-Associated Virus Vector Genomes. *J. Virol.* **84,** 8673–8682 (2010).

409. Pillay, S. *et al.* An essential receptor for adeno-associated virus infection. *Nature* **530,** 108–112 (2016).

410. Ding, W., Zhang, L., Yan, Z. & Engelhardt, J. F. Intracellular trafficking of adeno-associated viral vectors. *Gene Ther.* **12,** 873–880 (2005).

411. Wu, Z., Asokan, A. & Samulski, R. J. Adeno-associated Virus Serotypes: Vector Toolkit for Human Gene Therapy. *Mol. Ther.* **14,** 316–327 (2006).

412. Srivastava, A. In vivo tissue-tropism of adeno-associated viral vectors. *Curr. Opin. Virol.* **21,** 75–80 (2016).

413. Wu, Z., Yang, H. & Colosi, P. Effect of Genome Size on AAV Vector Packaging. *Mol. Ther.* **18,** 80–86 (2010).

414. Yan, Z., Zhang, Y., Duan, D. & Engelhardt, J. F. Trans-splicing vectors expand the utility of adeno-associated virus for gene therapy. *Proc. Natl. Acad. Sci.* **97,** 6716–6721 (2000).

415. McCarty, D. M., Monahan, P. E. & Samulski, R. J. Self-complementary recombinant adeno-associated

virus (scAAV) vectors promote efficient transduction independently of DNA synthesis. *Gene Ther.* **8,** 1248–1254 (2001).

416. Ferrari, F. K., Samulski, T., Shenk, T. & Samulski, R. J. Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. *J. Virol.* **70,** 3227–34 (1996).

417. Fisher, K. J. *et al.* Transduction with recombinant adeno-associated virus for gene therapy is limited by leading-strand synthesis. *J. Virol.* **70,** 520–32 (1996).

418. Li, M. *et al.* High-Efficiency Transduction of Fibroblasts and Mesenchymal Stem Cells by Tyrosine-Mutant AAV2 Vectors for Their Potential Use in Cellular Therapy. *Hum. Gene Ther.* **21,** 1527–1543 (2010).

419. Grimm, D. *et al.* In Vitro and In Vivo Gene Therapy Vector Evolution via Multispecies Interbreeding and Retargeting of Adeno-Associated Viruses. *J. Virol.* **82,** 5887–5911 (2008).

420. Zinn, E. *et al.* In Silico Reconstruction of the Viral Evolutionary Lineage Yields a Potent Gene Therapy Vector. *Cell Rep.* **12,** 1056–1068 (2015).

421. Schnepp, B. C., Clark, K. R., Klemanski, D. L., Pacak, C. A. & Johnson, P. R. Genetic fate of recombinant adeno-associated virus vector genomes in muscle. *J. Virol.* **77,** 3495–504 (2003).

422. Miller, D. G., Petek, L. M. & Russell, D. W. Adeno-associated virus vectors integrate at chromosome breakage sites. *Nat. Genet.* **36,** 767–773 (2004).

423. Honaramooz, A. *et al.* Adeno-associated virus (AAV)-mediated transduction of male germ line stem cells results in transgene transmission after germ cell transplantation. *FASEB J.* **22,** 374–82 (2008).

424. Russell, D. W. AAV loves an active genome. *Nat. Genet.* **34,** 241–242 (2003).

425. Nakai, H. *et al.* AAV serotype 2 vectors preferentially integrate into active genes in mice. *Nat. Genet.* **34,** 297–302 (2003).

426. Cervelli, T. *et al.* Processing of recombinant AAV genomes occurs in specific nuclear structures that overlap with foci of DNA-damage-response proteins. *J. Cell Sci.* **121,** 349–357 (2008).

427. Lamarche, B. J., Orazio, N. I. & Weitzman, M. D. The MRN complex in double-strand break repair and telomere maintenance. *FEBS Lett.* **584,** 3682–3695 (2010).

428. McCarty, D. M., Young, S. M. & Samulski, R. J. Integration of Adeno-Associated Virus (AAV) and Recombinant AAV Vectors. *Annu. Rev. Genet.* **38,** 819–845 (2004).

429. Mojica, F. J. M. & Rodriguez-Valera, F. The discovery of CRISPR in archaea and bacteria. *FEBS J.* **283,** 3162–3169 (2016).

430. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J. Bacteriol.* **169,** 5429–5433 (1987).

431. Nakata, A., Amemura, M. & Makino, K. Unusual nucleotide arrangement with repeated sequences in the Escherichia coli K-12 chromosome. *J. Bacteriol.* **171,** 3553–6 (1989).

432. Hermans, P. W. *et al.* Insertion element IS987 from Mycobacterium bovis BCG is located in a hot-spot integration region for insertion elements in Mycobacterium tuberculosis complex strains. *Infect. Immun.* **59,** 2695–705 (1991).

433. Mojica, F. J. M., Juez, G. & Rodriguez-Valera, F. Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. *Mol. Microbiol.* **9,** 613–621 (1993).

434. Jansen, R., Embden, J. D. A. van, Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43,** 1565–1575 (2002).

435. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *J. Mol. Evol.* **60,** 174–182 (2005).

436. Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science (80-.).* **315,** 1709–1712 (2007).

437. Brouns, S. J. J. *et al.* Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science (80-.).* **321,** 960–964 (2008).

438. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468,** 67–71 (2010).

439. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471,** 602–607 (2011).

440. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.* **13,** 722–736 (2015).

441. Shmakov, S. *et al.* Diversity and evolution of class 2 CRISPR–Cas systems. *Nat. Rev. Microbiol.* **15,** 169–182 (2017).

442. Nuñez, J. K., Lee, A. S. Y., Engelman, A. & Doudna, J. A. Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. *Nature* **519,** 193–198 (2015).

443. Hynes, A. P., Villion, M. & Moineau, S. Adaptation in bacterial CRISPR-Cas immunity can be driven by defective phages. *Nat. Commun.* **5,** 1–6 (2014).

444. Zetsche, B. *et al.* Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163,** 759–771 (2015).

445. Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A. & Charpentier, E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **532,** 517–521 (2016).

446. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. a. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507,** 62–67 (2014).

447. Sternberg, S. H., Lafrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **527,** 110–113 (2015).

448. Kim, H. & Kim, J.-S. A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.* **15,** 321–334 (2014).

449. Smih, F., Rouet, P., Romanienko, P. J. & Jasin, M. Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic Acids Res.* **23,** 5012–5019 (1995).

450. Shukla, V. K. *et al.* Precise genome modification in the crop species Zea mays using zinc-finger nucleases. *Nature* **459,** 437–441 (2009).

451. Bogdanove, A. J. & Voytas, D. F. TAL Effectors: Customizable Proteins for DNA Targeting. *Science (80-.).* **333,** 1843–1846 (2011).

452. Miller, J. C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29,** 143–148 (2011).

453. Zhang, F. *et al.* Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat. Biotechnol.* **29,** 149–153 (2011).

454. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science (80-.).* **339,** 823–826 (2013).

455. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science (80-.).* **339,** 819–823 (2013).

456. Ran, F. A. *et al.* Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* **154,** 1380–1389 (2013).

457. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31,** 822–826 (2013).

458. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31,** 827–832 (2013).

459. Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. *Nat. Biotechnol.* **32,** 1262–1267 (2014).

460. Nishimasu, H. *et al.* Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **156,** 935–949 (2014).

461. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513,** 569–573 (2014).

462. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science (80-.).* **351,** 84–88 (2016).

463. Kleinstiver, B. P. *et al.* High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529,** 490–495 (2016).

464. Chen, J. S. *et al.* Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature* **550,** 407–410 (2017).

465. Casini, A. *et al.* A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat. Biotechnol.* (2018). doi:10.1038/nbt.4066

466. Leenay, R. T. *et al.* Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. *Mol. Cell* **62,** 137–147 (2016).

467. Siksnys, V. & Gasiunas, G. Rewiring Cas9 to Target New PAM Sequences. *Mol. Cell* **61,** 793–794 (2016).

468. Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **168,** 20–36 (2017).

469. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523,** 481–485 (2015).

470. Anders, C., Bargsten, K. & Jinek, M. Structural Plasticity of PAM Recognition by Engineered Variants of the RNA-Guided Endonuclease Cas9. *Mol. Cell* **61,** 895–902 (2016).

471. Hirano, S., Nishimasu, H., Ishitani, R. & Nureki, O. Structural Basis for the Altered PAM Specificities of Engineered CRISPR-Cas9. *Mol. Cell* **61,** 886–894 (2016).

472. Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* 1–24 (2018). doi:10.1038/nature26155

473. Mali, P., Esvelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. *Nat. Methods* **10,** 957–963 (2013).

474. Morgan, S. L. *et al.* Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. *Nat. Commun.* **8,** 15993 (2017).

475. Hao, N., Shearwin, K. E. & Dodd, I. B. Programmable DNA looping using engineered bivalent dCas9 complexes. *Nat. Commun.* **8,** 1628 (2017).

476. Chen, B. *et al.* Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell* **155,** 1479–1491 (2013).

477. Ma, H. *et al.* Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc. Natl. Acad. Sci.* **112,** 3002–3007 (2015).

478. Ma, H. *et al.* Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nat. Biotechnol.* **34,** 528–530 (2016).

479. Fujita, T. & Fujii, H. Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. *Biochem. Biophys. Res. Commun.* **439,** 132–136 (2013).

480. Waldrip, Z. J. *et al.* A CRISPR-based approach for proteomic analysis of a single genomic locus. *Epigenetics* **9,** 1207–1211 (2014).

481. Liu, X. *et al.* In Situ Capture of Chromatin Interactions by Biotinylated dCas9. *Cell* **170,** 1028–1043.e19 (2017).

482. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533,** 420–424 (2016).

483. Gaudelli, N. M. *et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551,** 464–471 (2017).

484. Kim, Y. B. *et al.* Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* **35,** 371–376 (2017).

485. Hou, Z. *et al.* Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. *Proc. Natl. Acad. Sci.* **110,** 15644–15649 (2013).

486. Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520,** 186–191 (2015).

487. Esvelt, K. M. *et al.* Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods* **10,** 1116–1121 (2013).

488. Harrington, L. B. *et al.* A thermostable Cas9 with increased lifetime in human plasma. *Nat. Commun.* **8,** 1424 (2017).

489. Engler, C., Kandzia, R. & Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLoS One* **3,** e3647 (2008).

490. Kabadi, A. M., Ousterout, D. G., Hilton, I. B. & Gersbach, C. a. Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. *Nucleic Acids Res.* **42,** e147–e147 (2014).

491. Sakuma, T., Nishikawa, A., Kume, S., Chayama, K. & Yamamoto, T. Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Sci. Rep.* **4,** 5400 (2015).

492. Vad-Nielsen, J., Lin, L., Bolund, L., Nielsen, A. L. & Luo, Y. Golden Gate Assembly of CRISPR gRNA expression array for simultaneously targeting multiple genes. *Cell. Mol. Life Sci.* **73,** 4315–4325 (2016).

493. Nissim, L., Perli, S. D., Fridkin, A., Perez-Pinera, P. & Lu, T. K. Multiplexed and Programmable Regulation of Gene Networks with an Integrated RNA and CRISPR/Cas Toolkit in Human Cells. *Mol. Cell* **54,** 698–710 (2014).

494. Mefferd, A. L., Kornepati, A. V. R., Bogerd, H. P., Kennedy, E. M. & Cullen, B. R. Expression of CRISPR/Cas single guide RNAs using small tRNA promoters. *RNA* **21,** 1683–1689 (2015).

495. Xie, K., Minkenberg, B. & Yang, Y. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl. Acad. Sci.* **112,** 3570–3575 (2015).

496. Port, F. & Bullock, S. L. Augmenting CRISPR applications in Drosophila with tRNA-flanked sgRNAs. *Nat. Methods* **13,** 852–854 (2016).

497. Xu, L., Zhao, L., Gao, Y., Xu, J. & Han, R. Empower multiplex cell and tissue-specific CRISPR-mediated gene manipulation with self-cleaving ribozymes and tRNA. *Nucleic Acids Res.* gkw1048 (2016). doi:10.1093/nar/gkw1048

498. He, Y. *et al.* Self-cleaving ribozymes enable the production of guide RNAs from unlimited choices of promoters for CRISPR/Cas9 mediated genome editing. *J. Genet. Genomics* **44,** 469–472 (2017).

499. Yan, Q. *et al.* Multiplex CRISPR/Cas9-based genome engineering enhanced by Drosha-mediated sgRNA-shRNA structure. *Sci. Rep.* **6,** 38970 (2016).

500. Xie, C. *et al.* SgRNA Expression of CRIPSR-Cas9 System Based on MiRNA Polycistrons as a Versatile Tool to Manipulate Multiple and Tissue-Specific Genome Editing. *Sci. Rep.* **7,** 5795 (2017).

501. Zetsche, B. *et al.* Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nat. Biotechnol.* **35,** 31–34 (2016).

502. Wang, M., Mao, Y., Lu, Y., Tao, X. & Zhu, J. Multiplex Gene Editing in Rice Using the CRISPR-Cpf1 System. *Mol. Plant* **10,** 1011–1013 (2017).

503. Zhang, X. *et al.* Multiplex gene regulation by CRISPR-ddCpf1. *Cell Discov.* **3,** 17018 (2017).

504. Tak, Y. E. *et al.* Inducible and multiplex gene regulation using CRISPR–Cpf1-based transcription factors. *Nat. Methods* **14,** 1163–1166 (2017).

505. Kiani, S. *et al.* Cas9 gRNA engineering for genome editing, activation and repression. *Nat. Methods* **12,** 1051–1054 (2015).

506. Dahlman, J. E. *et al.* Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. *Nat. Biotechnol.* **33,** 1159–1161 (2015).

507. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152,** 1173–1183 (2013).

508. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* **41,** 7429–7437 (2013).

509. Gilbert, L. A. *et al.* CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell* **154,** 442–451 (2013).

510. Maeder, M. L. *et al.* CRISPR RNA–guided activation of endogenous human genes. *Nat. Methods* **10,** 977–979 (2013).

511. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9–based transcription factors. *Nat. Methods* **10,** 973–976 (2013).

512. Cheng, A. W. *et al.* Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.* **23,** 1163–1171 (2013).

513. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31,** 833–838 (2013).

514. Thakore, P. I., Black, J. B., Hilton, I. B. & Gersbach, C. A. Editing the epigenome: technologies for programmable transcription and epigenetic modulation. *Nat. Methods* **13,** 127–137 (2016).

515. Baron, U., Gossen, M. & Bujard, H. Tetracycline-controlled transcription in eukaryotes: novel transactivators with graded transactivation potential. *Nucleic Acids Res.* **25,** 2723–9 (1997).

516. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. *Cell* **159,** 635–646 (2014).

517. Kamiyama, D. *et al.* Versatile protein tagging in cells with split fluorescent protein. *Nat. Commun.* **7,** 11046 (2016).

518. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517,** 583–588 (2014).

519. Shechner, D. M., Hacisuleyman, E., Younger, S. T. & Rinn, J. L. Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nat. Methods* **12,** 664–670 (2015).

520. Maji, B. *et al.* Multidimensional chemical control of CRISPR–Cas9. *Nat. Chem. Biol.* **13,** 9–11 (2016).

521. Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* **12,** 326–328 (2015).

522. Chavez, A. *et al.* Comparison of Cas9 activators in multiple species. *Nat. Methods* **13,** 563–567 (2016).

523. Hu, J. H., Davis, K. M. & Liu, D. R. Chemical Biology Approaches to Genome Editing: Understanding, Controlling, and Delivering Programmable Nucleases. *Cell Chem. Biol.* **23,** 57–73 (2016).

524. González, F. *et al.* An iCRISPR Platform for Rapid, Multiplexable, and Inducible Genome Editing in Human Pluripotent Stem Cells. *Cell Stem Cell* **15,** 215–226 (2014).

525. Guo, J. *et al.* An inducible CRISPR-ON system for controllable gene activation in human pluripotent stem cells. *Protein Cell* **8,** 379–393 (2017).

526. Ferry, Q. R. V., Lyutova, R. & Fulga, T. A. Rational design of inducible CRISPR guide RNAs for de novo assembly of transcriptional programs. *Nat. Commun.* **8,** 14633 (2017).

527. Liu, Y. *et al.* Directing cellular information flow via CRISPR signal conductors. *Nat. Methods* **13,** 938–944 (2016).

528. Tang, W., Hu, J. H. & Liu, D. R. Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. *Nat. Commun.* **8,** 15939 (2017).

529. Liu, Y. *et al.* Engineering cell signaling using tunable CRISPR–Cpf1-based transcription factors. *Nat. Commun.* **8,** 2095 (2017).

530. Richter, F. *et al.* Switchable Cas9. *Curr. Opin. Biotechnol.* **48,** 119–126 (2017).

531. Zetsche, B., Volz, S. E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat. Biotechnol.* **33,** 139–142 (2015).

532. Nihongaki, Y., Kawano, F., Nakajima, T. & Sato, M. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nat. Biotechnol.* **33,** 755–760 (2015).

533. Gao, Y. *et al.* Complex transcriptional modulation with orthogonal and inducible dCas9 regulators. *Nat. Methods* **13,** 1043–1049 (2016).

534. Bao, Z., Jain, S., Jaroenpuntaruk, V. & Zhao, H. Orthogonal Genetic Regulation in Human Cells Using Chemically Induced CRISPR/Cas9 Activators. *ACS Synth. Biol.* **6,** 686–693 (2017).

535. Braun, S. M. G. *et al.* Rapid and reversible epigenome editing by endogenous chromatin regulators. *Nat. Commun.* **8,** 560 (2017).

536. Nihongaki, Y., Yamamoto, S., Kawano, F., Suzuki, H. & Sato, M. CRISPR-Cas9-based Photoactivatable Transcription System. *Chem. Biol.* **22,** 169–174 (2015).

537. Polstein, L. R. & Gersbach, C. A. A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nat. Chem. Biol.* **11,** 198–200 (2015).

538. Kleinjan, D. A., Wardrope, C., Nga Sou, S. & Rosser, S. J. Drug-tunable multidimensional synthetic gene control using inducible degron-tagged dCas9 effectors. *Nat. Commun.* **8,** 1191 (2017).

539. Senturk, S. *et al.* Rapid and tunable method to temporally control gene editing based on conditional Cas9 stabilization. *Nat. Commun.* **8,** 14370 (2017).

540. Oakes, B. L. *et al.* Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. *Nat. Biotechnol.* **34,** 646–651 (2016).

541. Davis, K. M., Pattanayak, V., Thompson, D. B., Zuris, J. A. & Liu, D. R. Small molecule–triggered Cas9 protein with improved genome-editing specificity. *Nat. Chem. Biol.* **11,** 316–318 (2015).

542. Zhou, X. X. *et al.* A single-chain photoswitchable CRISPR-Cas9 architecture for light-inducible gene editing and transcription. *ACS Chem. Biol.* acschembio.7b00603 (2017). doi:10.1021/acschembio.7b00603

543. Rose, J. C., Stephany, J. J., Wei, C. T., Fowler, D. M. & Maly, D. J. Rheostatic Control of Cas9-Mediated DNA Double Strand Break (DSB) Generation and Genome Editing. *ACS Chem. Biol.* acschembio.7b00652 (2017). doi:10.1021/acschembio.7b00652

544. Nguyen, D. P. *et al.* Ligand-binding domains of nuclear receptors facilitate tight control of split CRISPR activity. *Nat. Commun.* **7,** 12009 (2016).

545. Liu, K. I. *et al.* A chemical-inducible CRISPR–Cas9 system for rapid control of genome editing. *Nat. Chem. Biol.* **12,** 980–987 (2016).

546. Ishida, K. *et al.* Site-specific randomization of the endogenous genome by a regulatable CRISPR-Cas9 piggyBac system in human cells. *Sci. Rep.* **8,** 310 (2018).

547. Schwarz, K. A., Daringer, N. M., Dolberg, T. B. & Leonard, J. N. Rewiring human cellular input–output using modular extracellular sensors. *Nat. Chem. Biol.* **13,** 202–209 (2016).

548. Baeumler, T. A., Ahmed, A. A. & Fulga, T. A. Engineering Synthetic Signaling Pathways with Programmable dCas9-Based Chimeric Receptors. *Cell Rep.* **20,** 2639–2653 (2017).

549. Kipniss, N. H. *et al.* Engineering cell sensing and responses using a GPCR-coupled CRISPR-Cas system. *Nat. Commun.* **8,** 2212 (2017).

550. Hilton, I. B. *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* **33,** 510–517 (2015).

551. Kwon, D. Y., Zhao, Y.-T., Lamonica, J. M. & Zhou, Z. Locus-specific histone deacetylation using a synthetic CRISPR-Cas9-based HDAC. *Nat. Commun.* **8,** 15315 (2017).

552. Kearns, N. a *et al.* Functional annotation of native enhancers with a Cas9–histone demethylase fusion. *Nat. Methods* **12,** 401–403 (2015).

553. Vojta, A. *et al.* Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res.* **44,** 5615–5628 (2016).

554. Liu, X. S. *et al.* Editing DNA Methylation in the Mammalian Genome. *Cell* **167,** 233–247.e17 (2016).

555. Morita, S. *et al.* Targeted DNA demethylation in vivo using dCas9–peptide repeat and scFv–TET1 catalytic domain fusions. *Nat. Biotechnol.* **34,** 1060–1065 (2016).

556. Barkal, A. A., Srinivasan, S., Hashimoto, T., Gifford, D. K. & Sherwood, R. I. Cas9 Functionally Opens Chromatin. *PLoS One* **11,** e0152683 (2016).

557. Chen, F. *et al.* Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. *Nat. Commun.* **8,** 14958 (2017).

558. Kearns, N. a. *et al.* Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. *Development* **141,** 219–223 (2014).

559. Furuhata, Y., Nihongaki, Y., Sato, M. & Yoshimoto, K. Control of Adipogenic Differentiation in Mesenchymal Stem Cells via Endogenous Gene Activation Using CRISPR-Cas9. *ACS Synth. Biol.* **6,** 2191–2197 (2017).

560. Chakraborty, S. *et al.* A CRISPR/Cas9-Based System for Reprogramming Cell Lineage Specification. *Stem Cell Reports* **3,** 940–947 (2014).

561. Black, J. B. *et al.* Targeted Epigenetic Remodeling of Endogenous Loci by CRISPR/Cas9-Based Transcriptional Activators Directly Converts Fibroblasts to Neuronal Cells. *Cell Stem Cell* **19,** 406–414 (2016).

562. Liao, H.-K. *et al.* In Vivo Target Gene Activation via CRISPR/Cas9-Mediated Trans -epigenetic Modulation. *Cell* **171,** 1495–1507.e15 (2017).

563. Xiong, K. *et al.* RNA-Guided Activation of Pluripotency Genes in Human Fibroblasts. *Cell. Reprogram.* **19,** 189–198 (2017).

564. Gao, X. *et al.* Comparison of TALE designer transcription factors and the CRISPR/dCas9 in regulation of gene expression by targeting enhancers. *Nucleic Acids Res.* **42,** e155–e155 (2014).

565. Liu, P., Chen, M., Liu, Y., Qi, L. S. & Ding, S. CRISPR-Based Chromatin Remodeling of the Endogenous Oct4 or Sox2 Locus Enables Reprogramming to Pluripotency. *Cell Stem Cell* **0,** 1–10 (2018).

566. Garcia-Perez, J. L., Widmann, T. J. & Adams, I. R. The impact of transposable elements on mammalian development. *Development* **143,** 4101–4114 (2016).

567. Imbeault, M., Helleboid, P.-Y. & Trono, D. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **543,** 550–554 (2017).

568. Long, H. K., Prescott, S. L. & Wysocka, J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **167,** 1170–1187 (2016).

569. Kunarso, G. *et al.* Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.* **42,** 631–634 (2010).

570. Mi, S. *et al.* Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403,** 785–789 (2000).

571. Grow, E. J. *et al.* Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* **522,** 221–225 (2015).

572. Jachowicz, J. W. *et al.* LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat. Genet.* **49,** 1502–1510 (2017).

573. Wu, J. *et al.* The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534,** 652–657 (2016).

574. Göke, J. *et al.* Dynamic Transcription of Distinct Classes of Endogenous Retroviral Elements Marks Specific Populations of Early Human Embryonic Cells. *Cell Stem Cell* **16,** 135–141 (2015).

575. Theunissen, T. W. *et al.* Molecular Criteria for Defining the Naive Human Pluripotent State. *Cell Stem Cell* **19,** 502–515 (2016).

576. Hendrickson, P. G. *et al.* Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* **49,** 925–934 (2017).

577. Friedli, M. *et al.* Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency. *Genome Res.* **24,** 1251–1259 (2014).

578. Klawitter, S. *et al.* Reprogramming triggers endogenous L1 and Alu retrotransposition in human induced pluripotent stem cells. *Nat. Commun.* **7,** 10286 (2016).

579. Houck, C. M., Rinehart, F. P. & Schmid, C. W. A ubiquitous family of repeated DNA sequences in the human genome. *J. Mol. Biol.* **132,** 289–306 (1979).

580. Ullu, E. & Tschudi, C. Alu sequences are processed 7SL RNA genes. *Nature* **312,** 171–172 (1984).

581. Kramerov, D. A. & Vassetzky, N. S. Origin and evolution of SINEs in eukaryotic genomes. *Heredity (Edinb).* **107,** 487–495 (2011).

582. Quentin, Y. Origin of the Alu family: a family of Alu-like monomers gave birth to the left and the right arms of the Alu elements. *Nucleic Acids Res.* **20,** 3397–401 (1992).

583. Batzer, M. A. & Deininger, P. L. Alu repeats and human genomic diversity. *Nat. Rev. Genet.* **3,** 370–9 (2002).

584. Häsler, J. & Strub, K. Alu elements as regulators of gene expression. *Nucleic Acids Res.* **34,** 5491–5497 (2006).

585. Tsirigos, A. & Rigoutsos, I. Alu and B1 Repeats Have Been Selectively Retained in the Upstream and Intronic Regions of Genes of Specific Functional Classes. *PLoS Comput. Biol.* **5,** e1000610 (2009).

586. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).

587. Willoughby, D. A., Vilalta, A. & Oshima, R. G. An Alu Element from the K18 Gene Confers Position-independent Expression in Transgenic Mice. *J. Biol. Chem.* **275,** 759–768 (2000).

588. Roman, A. C. *et al.* Dioxin receptor and SLUG transcription factors regulate the insulator activity of B1 SINE retrotransposons via an RNA polymerase switch. *Genome Res.* **21,** 422–432 (2011).

589. Su, M., Han, D., Boyd-Kirkup, J., Yu, X. & Han, J.-D. J. Evolution of Alu Elements toward Enhancers. *Cell Rep.* **7,** 376–385 (2014).

590. Crepaldi, L. *et al.* Binding of TFIIIC to SINE Elements Controls the Relocation of Activity-Dependent Neuronal Genes to Transcription Factories. *PLoS Genet.* **9,** e1003699 (2013).

591. Policarpi, C. *et al.* Enhancer SINEs Link Pol III to Pol II Transcription in Neurons. *Cell Rep.* **21,** 2879–2894 (2017).

592. Polak, P. & Domany, E. Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. *BMC Genomics* **7,** 133 (2006).

593. Luo, Y., Lu, X. & Xie, H. Dynamic Alu Methylation during Normal Development, Aging, and Tumorigenesis. *Biomed Res. Int.* **2014,** 1–12 (2014).

594. Oei, S.-L. *et al.* Clusters of regulatory signals for RNA polymerase II transcription associated with Alu family repeats and CpG islands in human promoters. *Genomics* **83,** 873–882 (2004).

595. Töhönen, V. *et al.* Novel PRD-like homeodomain transcription factors and retrotransposon elements in early human development. *Nat. Commun.* **6,** 8207 (2015).

596. Ge, S. X. Exploratory bioinformatics investigation reveals importance of "junk" DNA in early embryo development. *BMC Genomics* **18,** 200 (2017).

597. Jouhilahti, E.-M. *et al.* The human PRD-like homeobox gene LEUTX has a central role in embryo genome activation. *Development* **143,** 3459–3469 (2016).

598. Madissoon, E. *et al.* Characterization and target genes of nine human PRD-like homeobox domain genes expressed exclusively in early embryos. *Sci. Rep.* **6,** 28995 (2016).

599. Bolotin, E. *et al.* Nuclear Receptor HNF4α Binding Sequences are Widespread in Alu Repeats. *BMC Genomics* **12,** 560 (2011).

600. Morales-Hernández, A. *et al.* Alu retrotransposons promote differentiation of human carcinoma cells through the aryl hydrocarbon receptor. *Nucleic Acids Res.* **44,** 4665–4683 (2016).

601. Vansant, G. & Reynolds, W. F. The consensus sequence of a major Alu subfamily contains a functional retinoic acid response element. *Proc. Natl. Acad. Sci.* **92,** 8229–8233 (1995).

602. Zolotukhin, S. *et al.* Recombinant adeno-associated virus purification using novel methods improves infectious titer and yield. *Gene Ther.* **6,** 973–985 (1999).

603. Toivonen, S. *et al.* Activin A and Wnt-dependent specification of human definitive endoderm cells. *Exp. Cell Res.* **319,** 2535–2544 (2013).

604. Cermak, T. *et al.* Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* **39,** 7879–7879 (2011).

605. Qin, J. Y. *et al.* Systematic Comparison of Constitutive Promoters and the Doxycycline-Inducible Promoter. *PLoS One* **5,** e10611 (2010).

606. Chen, M.-J. *et al.* Reprogramming Adipose Tissue-Derived Mesenchymal Stem Cells into Pluripotent Stem Cells by a Mutant Adeno-Associated Viral Vector. *Hum. Gene Ther. Methods* **25,** 72–82 (2014).

607. Meilinger, D. *et al.* Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells. *EMBO Rep.* **10,** 1259–1264 (2009).

608. Lo Sardo, V. *et al.* Influence of donor age on induced pluripotent stem cells. *Nat. Biotechnol.* **35,** 69–74 (2016).

609. Choudhury, A. R. *et al.* Cdkn1a deletion improves stem cell function and lifespan of mice with dysfunctional telomeres without accelerating cancer formation. *Nat. Genet.* **39,** 99–105 (2007).

610. Brown, J. P. Bypass of Senescence After Disruption of p21CIP1/WAF1 Gene in Normal Diploid Human Fibroblasts. *Science (80-.).* **277,** 831–834 (1997).

611. Jiang, J. *et al.* Zscan4 promotes genomic stability during reprogramming and dramatically improves the quality of iPS cells as demonstrated by tetraploid complementation. *Cell Res.* **23,** 92–106 (2013).

612. Le, R. *et al.* Enhanced Telomere Rejuvenation in Pluripotent Cells Reprogrammed via Nuclear Transfer Relative to Induced Pluripotent Stem Cells. *Cell Stem Cell* **14,** 27–39 (2014).

613. Yamanaka, S. & Blau, H. M. Nuclear reprogramming to a pluripotent state by three approaches. *Nature* **465,** 704–712 (2010).

614. Moberg, K. H., Tyndall, W. A. & Hall, D. J. Wild-type murine p53 represses transcription from the murinec-myc promoter in a human glial cell line. *J. Cell. Biochem.* **49,** 208–215 (1992).

615. Ho, J. S. L., Ma, W., Mao, D. Y. L. & Benchimol, S. p53-Dependent Transcriptional Repression of c-myc Is Required for G1 Cell Cycle Arrest. *Mol. Cell. Biol.* **25,** 7423–7431 (2005).

616. Iwamoto, M., Björklund, T., Lundberg, C., Kirik, D. & Wandless, T. J. A general chemical method to regulate protein stability in the mammalian central nervous system. *Chem. Biol.* **17,** 981–8 (2010).

617. Van Bortle, K. & Corces, V. G. tDNA insulators and the emerging role of TFIIIC in genome organization. *Transcription* **3,** 277–284 (2012).

618. Yuen, K. C., Slaughter, B. D. & Gerton, J. L. Condensin II is anchored by TFIIIC and H3K4me3 in the mammalian genome and supports the expression of active dense gene clusters. *Sci. Adv.* **3,** e1700191 (2017).

13. Original Publications